

Portland State University

**PDXScholar**

---

Dissertations and Theses

Dissertations and Theses

---

Winter 3-7-2017

# Computational Algorithms for Improved Representation of the Model Error Covariance in Weak-Constraint 4D-Var

Jeremy A. Shaw  
*Portland State University*

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/open\\_access\\_etds](https://pdxscholar.library.pdx.edu/open_access_etds)



Part of the [Mathematics Commons](#), and the [Numerical Analysis and Computation Commons](#)

**Let us know how access to this document benefits you.**

---

## Recommended Citation

Shaw, Jeremy A., "Computational Algorithms for Improved Representation of the Model Error Covariance in Weak-Constraint 4D-Var" (2017). *Dissertations and Theses*. Paper 3473.

<https://doi.org/10.15760/etd.5357>

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

Computational Algorithms for Improved Representation of the Model Error  
Covariance in Weak-Constraint 4D-Var

by  
Jeremy A. Shaw

A dissertation submitted in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy  
in  
Mathematical Sciences

Dissertation Committee:  
Dacian Daescu, Chair  
Bin Jiang  
Ian Dinwoodie  
Hamid Moradkhani  
Y. C. Jenq

Portland State University  
2017

© 2017 Jeremy A. Shaw

## Abstract

Four-dimensional variational data assimilation (4D-Var) provides an estimate to the state of a dynamical system through the minimization of a cost functional that measures the distance to a prior state (background) estimate and observations over a time window. The analysis fit to each information input component is determined by the specification of the error covariance matrices in the data assimilation system (DAS). Weak-constraint 4D-Var (w4D-Var) provides a theoretical framework to account for modeling errors in the analysis scheme. In addition to the specification of the background error covariance matrix, the w4D-Var formulation requires information on the model error statistics and specification of the model error covariance. Up to now, the increased computational cost associated with w4D-Var has prevented its practical implementation. Various simplifications to reduce the computational burden have been considered, including writing the model error covariance as a scalar multiple of the background error covariance and modeling the model error.

In this thesis, the main objective is the development of computationally feasible techniques for the improved representation of the model error statistics in a data assimilation system. Three new approaches are considered.

1. A Monte Carlo method that uses an ensemble of w4D-Var systems to obtain flow-dependent estimates to the model error statistics.
2. The evaluation of statistical diagnostic equations involving observation residuals to estimate the model error covariance matrix.

3. An adaptive tuning procedure based on the sensitivity of a short-range forecast error measure to the model error DAS parametrization.

The validity and benefits of these approaches are shown in two stages of numerical experiments. A proof-of-concept is shown using the Lorenz multi-scale model and the shallow water equations for a one-dimensional domain. The results show the potential of these methodologies to produce improved state estimates, as compared to other approaches in data assimilation. It is expected that the techniques presented will find an extended range of applications to assess and improve the performance of a w4D-Var system.

## Dedication

*This dissertation is dedicated to my parents, Susan and Marvin, who are always proud of me, to my identical twin brother, Jeff, who looks like me a little, and to my sister, Trina, who says she hates math. I also dedicate this work to Paul Collins and the rest of my friends and family for being so supportive.*

## Acknowledgments

I would like to thank my advisor, Prof. Dacian Daescu, for introducing me to the topic of data assimilation and for all of the time spent meeting with me to discuss my work.

Additionally, I thank all of the other professors that taught me along the way for the knowledge I have gained.

For showing me how to use the university's research computing systems, I thank Nichole Schimanski.

Lastly, I want to thank the anonymous reviewers for their suggestions on the version of Chapter 3 that appears in *Procedia Computer Science*.

## Table of Contents

<b>Abstract</b>	<b>i</b>
<b>Dedication</b>	<b>iii</b>
<b>Acknowledgments</b>	<b>iv</b>
<b>List of Tables</b>	<b>vii</b>
<b>List of Figures</b>	<b>viii</b>
<b>List of Algorithms</b>	<b>x</b>
<b>Chapter 1 Introduction</b>	<b>1</b>
1.1 Overview of Data Assimilation Methods . . . . .	1
1.2 The Tangent Linear and Adjoint Models . . . . .	7
1.3 Research Objectives . . . . .	9
1.4 Lorenz Multi-Scale Model . . . . .	10
1.5 Organization . . . . .	15
<b>Chapter 2 Weak-Constraint 4D-Var</b>	<b>16</b>
2.1 The Cost Functional . . . . .	16
2.2 The Incremental Algorithm . . . . .	19
2.3 The Analysis Equation . . . . .	23
<b>Chapter 3 Ensemble Data Assimilation</b>	<b>29</b>
3.1 Derivation of the Model Error Ensemble . . . . .	30
3.2 Ensemble-based w4D-Var Schemes . . . . .	32
3.3 Ensemble Error Statistics . . . . .	35
3.4 Numerical Results . . . . .	38
3.5 Conclusion . . . . .	43
<b>Chapter 4 Observation Space Diagnostics</b>	<b>45</b>
4.1 Derivation of Diagnostics . . . . .	46
4.2 A Special Case . . . . .	51
4.3 Numerical Results . . . . .	53



4.4	Conclusion . . . . .	56
<b>Chapter 5</b>	<b>Sensitivity Analysis</b>	<b>58</b>
5.1	Sensitivity Analysis . . . . .	59
5.2	Sensitivity Equations . . . . .	60
5.3	Sensitivities with Matrix Decomposition . . . . .	63
5.4	Sensitivities Using a Single Outer Loop Iteration . . . . .	67
5.5	Adaptive Tuning of Model Error Parameters . . . . .	69
5.6	Numerical Results . . . . .	71
5.7	Conclusion . . . . .	80
<b>Chapter 6</b>	<b>Experiments with the Shallow Water Equations</b>	<b>82</b>
6.1	The Discrete Model . . . . .	83
6.2	Data Assimilation System . . . . .	85
6.3	Numerical Results . . . . .	87
6.4	Conclusion . . . . .	92
<b>Chapter 7</b>	<b>Conclusions and Future Directions</b>	<b>93</b>
	<b>References</b>	<b>95</b>
<b>Appendix A</b>	<b>Ensemble Covariance Implementation</b>	<b>102</b>
<b>Appendix B</b>	<b>The Discrete Shallow Water Equations</b>	<b>104</b>
<b>Appendix C</b>	<b>Tangent Linear and Adjoint of the Discrete Shallow Water Model</b>	<b>107</b>

## List of Tables

Table 5.1	Summary of forecast error sensitivities to model error input parameters in w4D-Var. . . . .	68
Table 5.2	Specifications of $\mathbf{Q}_+$ and $\mathbf{Q}_-$ given a perturbation in $\mathbf{Q}$ or one of its matrix factors. . . . .	74
Table 5.3	Finite difference errors and log error ratios for each of the model error sensitivity equations. . . . .	76
Table 6.1	Finite difference errors and log error ratios for each of the model error sensitivity equations for the discrete shallow water model. .	87

## List of Figures

Figure 1.1	(a) Typical behavior of the LZ96 model, (b) The estimate of the true model error covariance matrix . . . . .	11
Figure 2.1	A demonstration of the incremental method showing the successive quadratic approximations of an example cost function. . . .	25
Figure 3.1	The ratio of the global monthly analysis errors to the control. . .	41
Figure 3.2	The ratio of the average oma differences to the control. . . . .	41
Figure 3.3	Comparison of the prescribed model error variance to the time-averaged ensemble and hybrid model error variances. . . . .	43
Figure 3.4	The ratio of the time- and space-averaged analysis errors to the control verses the hybrid scalar weight $\alpha$ for the model error covariances. . . . .	43
Figure 4.1	(a) Posterior estimate of $\mathbf{Q}^t$ , (b) Twentieth rows of $\mathbf{Q}^t$ and the posterior estimate. . . . .	54
Figure 4.2	Specification of the background error correlation matrix. . . . .	54
Figure 4.3	(a) Posterior estimate of $\mathbf{R}^t$ , (b) Twentieth rows of $\mathbf{R}^t$ and the posterior estimate. . . . .	55
Figure 5.1	A visual of the model error parameter tuning procedure. . . . .	70
Figure 5.2	(a) A visual of the rotating observing system, (b) The specification of the background error correlation matrix. . . . .	73
Figure 5.3	The ratio of the global monthly errors to the control w4D-Var experiment. . . . .	78
Figure 5.4	The step-size for the gradient descent update of the model error covariance matrices for each data assimilation cycle over a two-year period. . . . .	78
Figure 5.5	The same as Figure 5.3, but compares the monthly error ratios of the sensitivity DAS to the ensemble-based DASs. . . . .	79
Figure 5.6	The ratio of the analysis errors for the sensitivity and ensemble-based systems over each grid point. . . . .	79
Figure 5.7	The ratio of the analysis errors to the control for each time in the data assimilation window. . . . .	79

Figure 6.1	Schematic representation of the shallow water model. . . . .	83
Figure 6.2	Shallow water diagram with bottom topography given by (6.3). .	84
Figure 6.3	The “true” state of the shallow water model. . . . .	85
Figure 6.4	(a) The estimated model error covariance matrix for the discrete shallow water model, (b) Zoomed-in view of $\mathbf{Q}^t$ . . . . .	86
Figure 6.5	The ratio of the global analysis errors to the control w4D-Var experiment. . . . .	89
Figure 6.6	The step-sizes for the gradient descent update of the model error covariance matrices for each data assimilation cycle. . . . .	89
Figure 6.7	(a) Posterior estimate of $\mathbf{Q}^t$ , (b) Zoomed-in view of the $\mathbf{Q}$ -diagnostic matrix. . . . .	90
Figure 6.8	(a) Posterior estimate of $\mathbf{R}^t$ , (b) Zoomed-in view of the $\mathbf{R}$ -diagnostic matrix. . . . .	91

## List of Algorithms

Algorithm 3.1	Computation of the ensemble estimates of model error. . . . .	33
Algorithm 5.1	Procedure to update the model error covariance matrix using sensitivities. . . . .	72

## **Chapter 1**

### **Introduction**

Environmental processes can be realistically described by mathematical models of the system dynamics, which can be used to make predictions about future behavior of the system, provided that the initial conditions for the forecast are accurate [67]. However, it is generally the case that the data available incompletely describe all the states and data may contain random noise, causing a discrepancy between the true state of the dynamical system and a forecast of the true state [53]. The purpose of data assimilation is to incorporate all available information, measured observations and background knowledge of the state, with a mathematical model in order to estimate as accurately as possible the state of the system [74]. Once obtained, the estimate (analysis) can be used as input into another operation, for example, as the initial state for a numerical weather forecast [6].

#### **1.1 Overview of Data Assimilation Methods**

There are various types of data assimilation techniques used for research in a number of different applications, including, but not limited to, numerical weather prediction (NWP), oceanography, geosciences, hydrology, and signal processing. A brief review of several popular data assimilation methods is now presented.

### 1.1.1 The Kalman Filter

The Kalman filter [51, 52] is a recursive algorithm that uses noisy observations to produce an analysis of a linear dynamical system

$$\mathbf{x}_i^t = \mathbf{M}_i \mathbf{x}_{i-1}^t + \boldsymbol{\varepsilon}_i^q \quad (1.1)$$

where  $\mathbf{M}_i \in \mathbb{R}^{n \times n}$  is nonsingular and  $\boldsymbol{\varepsilon}_i^q$  denotes unbiased model error with covariance matrix  $\mathbf{Q}_i$ . An unbiased background estimate  $\mathbf{x}_0^b$  of  $\mathbf{x}_0^t$  is known with unbiased error  $\boldsymbol{\varepsilon}_0^b = \mathbf{x}_0^t - \mathbf{x}_0^b$  and error covariance matrix  $\mathbf{B}_0$ , which is uncorrelated with model error. Noisy observations satisfy

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{x}_i^t + \boldsymbol{\varepsilon}_i^o \quad (1.2)$$

where  $\mathbf{H}_i \in \mathbb{R}^{p \times n}$  represents the observation operator, mapping the state  $\mathbf{x}_i$  into observation space, as typically there are fewer observations than variables in the model, and  $\boldsymbol{\varepsilon}_i^o$  is the unbiased observation error with covariance matrix  $\mathbf{R}_i$ . With these assumptions, the Kalman filter finds an unbiased estimate  $\mathbf{x}_i^a$  of  $\mathbf{x}_i^t$  which is a linear function of  $\mathbf{x}_i^b$  and  $\mathbf{y}_i$  that minimizes the mean square error

$$\mathbb{E}[(\mathbf{x}_i^t - \mathbf{x}_i^a)^\top (\mathbf{x}_i^t - \mathbf{x}_i^a)]. \quad (1.3)$$

The solution to this problem is

$$\mathbf{x}_i^a = \mathbf{x}_i^b + \mathbf{K}_i [\mathbf{y}_i - \mathbf{H}_i \mathbf{x}_i^b] \quad (1.4)$$

where  $\mathbf{K}_i$  is the *Kalman gain* matrix

$$\mathbf{K}_i = \mathbf{B}_i \mathbf{H}_i^T [\mathbf{H}_i \mathbf{B}_i \mathbf{H}_i^T + \mathbf{R}_i]^{-1}. \quad (1.5)$$

In addition, the analysis error  $\boldsymbol{\varepsilon}_i^a = \mathbf{x}_i^t - \mathbf{x}_i^a$  has the covariance matrix

$$\mathbf{A}_i = [\mathbf{I}_n - \mathbf{K}_i \mathbf{H}_i] \mathbf{B}_i. \quad (1.6)$$

Once the analysis has been obtained, the model is used to forecast  $\mathbf{x}_i^a$ , which produces the background estimate for the next time-step, that is,  $\mathbf{x}_{i+1}^b = \mathbf{M}_i \mathbf{x}_i^a$ , whose error covariance matrix satisfies

$$\mathbf{B}_{i+1} = \mathbf{M}_i \mathbf{A}_i \mathbf{M}_i^T + \mathbf{Q}_i. \quad (1.7)$$

The nature of the Kalman filter makes it desirable to use. A famous application of the Kalman filter is its use in the Apollo moon project [39]. However, not all models are linear. In the event of a nonlinear dynamical system

$$\mathbf{x}_i^t = \mathcal{M}_i(\mathbf{x}_{i-1}^t) + \boldsymbol{\varepsilon}_i^q \quad (1.8)$$

and nonlinear observation operator  $\mathbf{h}_i$ , which gives the observations

$$\mathbf{y}_i = \mathbf{h}_i(\mathbf{x}_i^t) + \boldsymbol{\varepsilon}_i^o, \quad (1.9)$$

the Kalman filter does not apply without modification.



### 1.1.2 The Extended Kalman Filter

Due to the nonlinearity of the model  $\mathcal{M}_i$  and observation operator  $\mathbf{h}_i$ , their linearizations are necessary

$$\mathbf{h}_i(\mathbf{x}_i) \approx \mathbf{h}_i(\mathbf{x}_i^b) + \mathbf{H}_i^b(\mathbf{x}_i - \mathbf{x}_i^b) \quad (1.10)$$

$$\mathcal{M}_i(\mathbf{x}_{i-1}) \approx \mathcal{M}_i(\mathbf{x}_{i-1}^a) + \mathbf{M}_i^a(\mathbf{x}_{i-1} - \mathbf{x}_{i-1}^a), \quad (1.11)$$

where  $\mathbf{H}_i^b$  is the Jacobian of  $\mathbf{h}_i$  evaluated at  $\mathbf{x}_i^b$  and  $\mathbf{M}_i^a$  is the Jacobian of  $\mathcal{M}_i$  evaluated at  $\mathbf{x}_{i-1}^a$ . With these approximations, the Kalman filter equations are used to find the analysis state  $\mathbf{x}_i^a$  and use the linearized model  $\mathbf{M}_i^a$  to forecast for the next iteration. Using the Kalman filter equations for nonlinear  $\mathcal{M}_i$  and/or  $\mathbf{h}_i$  gives the extended Kalman filter (EKF). By using the linearizations, the Kalman filter equations are no longer exact, but only approximations, meaning what is called  $\mathbf{B}_i$  in the EKF is not truly the background error covariance. If the model  $\mathcal{M}_i$  and the observation operators  $\mathbf{h}_i$  are weakly nonlinear, then the approximation is a good one. The books by Anderson and Moore [2] and Jazwinski [50] provide a complete derivation of the EKF.

### 1.1.3 The Ensemble Kalman Filter

Evensen [31] introduces a Monte Carlo alternative, known as the ensemble Kalman filter (EnKF). The Kalman filter equations are used as in the EKF, however  $\mathbf{B}_i$  is replaced with an ensemble covariance matrix  $\mathbf{B}_e$ , which can be found at much lower computational cost than the cost of (1.7). Evensen [33] presents the mathematical framework of the EnKF and provides a historical overview of its development.

At each time-step,  $\mathbf{B}_e$  is formed from an ensemble of  $N_e$  members  $\mathbf{x}_j^b$  and  $j = 1, 2, \dots, N_e$  is the member index. The  $n \times N_e$  matrix  $\mathbf{X}$  is defined to be the rectangular

matrix whose columns are the scaled differences between the ensemble members and the ensemble mean  $\bar{\mathbf{x}}$  [12].

$$\mathbf{X} = \frac{1}{\sqrt{N_e - 1}} \begin{bmatrix} \mathbf{x}_1^b - \bar{\mathbf{x}} & \mathbf{x}_2^b - \bar{\mathbf{x}} & \cdots & \mathbf{x}_M^b - \bar{\mathbf{x}} \end{bmatrix} \quad (1.12)$$

The ensemble covariance  $\mathbf{B}_e$  is then given by

$$\mathbf{B}_e = \mathbf{X}\mathbf{X}^T. \quad (1.13)$$

Other developments in the EnKF occurred, and some of those results are summarized here. Evensen and van Leeuwen [34] re-derive the EnKF as a suboptimal solver for the general Bayesian problem. Hamill and Snyder [42] construct a hybrid data assimilation system by combining the EnKF and 3D-Var. Mitchell and Houtekamer [66] introduce an adaptive EnKF which accounts for the model error by estimating some of the model error statistics. Houtekamer and Mitchell [47] localize covariance matrices by using the Hadamard (Schur) product with a correlation function. Hamill et al. [43] investigate the relationship between ensemble size and distant covariances. Anderson [3] discusses different ensemble Kalman filters and provides an efficient two-step update procedure. Evensen [32] develops a new square root implementation of the EnKF.

#### 1.1.4 Four-Dimensional Variational Data Assimilation

In variational data assimilation, the problem of finding an analysis is formulated to be the minimizer of a cost functional, where optimization techniques can be applied. The four-dimensional variational data assimilation (4D-Var) cost functional measures the distance to a prior state (background) estimate and observations [56] over a time window  $[t_0, t_N]$ . The analysis fit to each information input component is determined

by the specification of the error covariance matrices in the data assimilation system (DAS). The works by Sasaki [71] and Talagrand and Courtier [75] are some early examples focusing on 4D-Var.

In a data assimilation window  $[t_0, t_N]$ , the analysis states  $\mathbf{x}_i^a$  depend only on the initial condition  $\mathbf{x}_0^a$ , which is obtained by minimizing the cost functional

$$J(\mathbf{x}_0) = \frac{1}{2}[\mathbf{x}_0 - \mathbf{x}_0^b]^T \mathbf{B}^{-1}[\mathbf{x}_0 - \mathbf{x}_0^b] + \frac{1}{2} \sum_{i=0}^N [\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)]^T \mathbf{R}_i^{-1}[\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)] \quad (1.14)$$

where the observations  $\mathbf{y}_i$  are valid at  $t_i \in [t_0, t_N]$  and the states  $\mathbf{x}_i$  at time  $t_i$  are related by the perfect model assumption

$$\mathbf{x}_i = \mathcal{M}_i(\mathbf{x}_{i-1}), \quad (1.15)$$

Using this model constraint,

$$\mathbf{x}_i^a = \mathcal{M}_i(\mathbf{x}_{i-1}^a). \quad (1.16)$$

The optimality condition satisfied by  $\mathbf{x}_0^a$  is  $\nabla_{\mathbf{x}_0} J(\mathbf{x}_0^a) = \mathbf{0}$ , or

$$\mathbf{B}^{-1}[\mathbf{x}_0 - \mathbf{x}_0^b] - \sum_{i=0}^N \mathbf{M}_{0 \rightarrow i}^T \mathbf{H}_i^T \mathbf{R}_i^{-1}[\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)] = \mathbf{0} \quad (1.17)$$

where  $\mathbf{M}_{0 \rightarrow i}$  is the Jacobian of the model integrated from time  $t_0$  to  $t_i$ . It is a special case of weak-constraint 4D-Var, in which the model relating the time-distributed states  $\mathbf{x}_i$  is not perfect and model error exists.

### 1.1.5 Weak-Constraint 4D-Var

Weak-constraint 4D-Var (w4D-Var) provides a theoretical framework to account for modeling errors in the analysis scheme. In addition to the specification of the background error covariance matrix, the w4D-Var formulation requires information on the model error statistics and specification of the model error covariance  $\mathbf{Q}$ . Instead of the perfect model assumption, the model equations

$$\mathbf{x}_i^t = \mathcal{M}_i(\mathbf{x}_{i-1}^t) + \boldsymbol{\varepsilon}_i^q \quad (1.18)$$

are imposed as a weak constraint of the optimization. The w4D-Var cost function is

$$\begin{aligned} J(\mathbf{x}_0, \dots, \mathbf{x}_N) = & \frac{1}{2}[\mathbf{x}_0 - \mathbf{x}_0^b]^T \mathbf{B}^{-1}[\mathbf{x}_0 - \mathbf{x}_0^b] + \frac{1}{2} \sum_{i=0}^N [\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)]^T \mathbf{R}_i^{-1} [\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)] \\ & + \frac{1}{2} \sum_{i=1}^N [\mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1}) - \mathbf{q}_i]^T \mathbf{Q}_i^{-1} [\mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1}) - \mathbf{q}_i], \end{aligned} \quad (1.19)$$

where  $\mathbf{q}_i = \mathbb{E}[\boldsymbol{\varepsilon}_i^q]$  is the model error bias vector. To help reduce the computational burden of implementing w4D-Var, several simplifications have been considered, including writing the model error covariance as a scalar multiple of the background error covariance (see [25] for example) and modeling the model error [41, 82, 83]. Further discussion of w4D-Var takes place in Chapter 2.

## 1.2 The Tangent Linear and Adjoint Models

An important component of variational data assimilation methods is the use of adjoint models in the computation of gradients. To a first-order approximation, the propagation of a vector  $\dot{\mathbf{x}}$  through a nonlinear model  $\mathcal{M}(\mathbf{x})$  is determined by the

tangent linear model

$$\dot{\mathbf{y}} = \mathbf{M}(\mathbf{x})\dot{\mathbf{x}}, \quad (1.20)$$

the product of the Jacobian matrix  $\mathbf{M}$  with  $\dot{\mathbf{x}}$ , where the dot is used to indicate a tangent linear variable. From linear algebra, it is known that for every linear operator  $\mathbf{M}$ , there exists an adjoint operator  $\mathbf{M}^T$  such that

$$\langle \mathbf{M}\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{M}^T\mathbf{y} \rangle. \quad (1.21)$$

The adjoint is used in the computation of the product of the Jacobian transpose  $\mathbf{M}^T(\mathbf{x})$  with a vector  $\bar{\mathbf{y}}$

$$\bar{\mathbf{x}} = \mathbf{M}^T(\mathbf{x})\bar{\mathbf{y}}, \quad (1.22)$$

where the bar is used to indicate adjoint variables. From the 4D-Var optimality condition (1.17), the need for adjoints is clear. In particular, the adjoint of the forecast model  $\mathbf{M}_{0 \rightarrow i}^T$ , which propagates the input backwards in time from  $t_i$  to  $t_0$ , and the adjoint of the observation operator  $\mathbf{H}_i^T$  are needed.

Implementation of the tangent linear and adjoint models involve the development of specialized code that produce the matrix-vector product without forming the Jacobian matrix. Giering and Kaminski [38] and Griewank and Walther [40] discuss the principles of tangent linear and adjoint code construction in detail. In essence, the tangent linear model is based on the concept that if  $\mathbf{y} = \mathcal{M}(\mathbf{x})$ , the corresponding

tangent linear statement for output  $y_i$  is given by

$$\dot{y}_i = \sum_j \frac{\partial y_i}{\partial x_j} \dot{x}_j, \quad (1.23)$$

involving the derivatives of  $y_i$  with respect to all of the independent variables. The adjoint statement for  $\mathbf{y} = \mathcal{M}(\mathbf{x})$  is based on

$$\bar{x}_i = \sum_j \frac{\partial y_j}{\partial x_i} \bar{y}_j, \quad (1.24)$$

involving the derivatives of all outputs  $y_j$  with respect to  $x_i$ . The derivation of tangent linear and adjoint code can be very time consuming and prone to errors when done by hand. Various tools have been developed for automatic adjoint code generation, such as Tapenade [44], which work on the principle that computer programs may be seen as a sequence of mathematical statements and function evaluations. Such computer programs can be differentiated by repeatedly applying the chain rule and the computed derivatives are accurate to machine precision. The tangent linear model is computed by propagating the derivatives of intermediate variables with respect to the inputs, referred to as forward mode. The adjoint model is evaluated in reverse mode by backward propagating the derivatives of the outputs with respect to the intermediate variables. An example of creating tangent linear and adjoint models is in Section 1.4.

### 1.3 Research Objectives

This dissertation aims to develop computationally feasible techniques for the improved representation of the model error statistics in a w4D-Var DAS. Since the analysis obtained from a w4D-Var DAS is determined by the input parameters, bet-

ter representation of the model error statistics will lead to a more accurate analysis. Three new approaches are considered for achieving this goal.

1. A Monte Carlo method that uses an ensemble of w4D-Var systems to obtain flow-dependent estimates to the model error statistics.
2. The evaluation of statistical diagnostic equations involving observation residuals to estimate the model error covariance matrix.
3. An adaptive tuning procedure based on the sensitivity (derivative) of a short-range forecast error measure to the model error DAS parametrization.

The theoretical foundations and development of these methodologies are separated into three chapters. Each of these chapters also contains a review of the pertinent published literature and preliminary results from the first stage of numerical experiments demonstrating the merits of the related content using the Lorenz multi-scale model, detailed next.

#### 1.4 Lorenz Multi-Scale Model

The first stage of numerical experiments involves the multi-scale model of Edward Lorenz [63]

$$\frac{d\hat{x}_k}{dt} = \hat{x}_{k-1}(\hat{x}_{k+1} - \hat{x}_{k-2}) - \hat{x}_k - \frac{hc}{b} \sum_{j=1}^J \hat{y}_{jk} + F \quad (1.25a)$$

$$\frac{d\hat{y}_{jk}}{dt} = cb\hat{y}_{j+1,k}(\hat{y}_{j-1,k} - \hat{y}_{j+2,k}) - c\hat{y}_{jk} + \frac{hc}{b}\hat{x}_k \quad (1.25b)$$

where  $k = 1, \dots, K$  and  $j = 1, \dots, J$ . The  $\hat{y}_{jk}$  variables vary at a smaller scale than the  $\hat{x}_k$  variables and are arranged as  $\hat{y}_{11}, \hat{y}_{21}, \dots, \hat{y}_{J1}, \hat{y}_{12}, \dots, \hat{y}_{J2}, \dots, \hat{y}_{JK}$ . They also extend cyclically so that  $\hat{y}_{J+1,1} = \hat{y}_{11}$ . The model given by (1.25) will be referred to

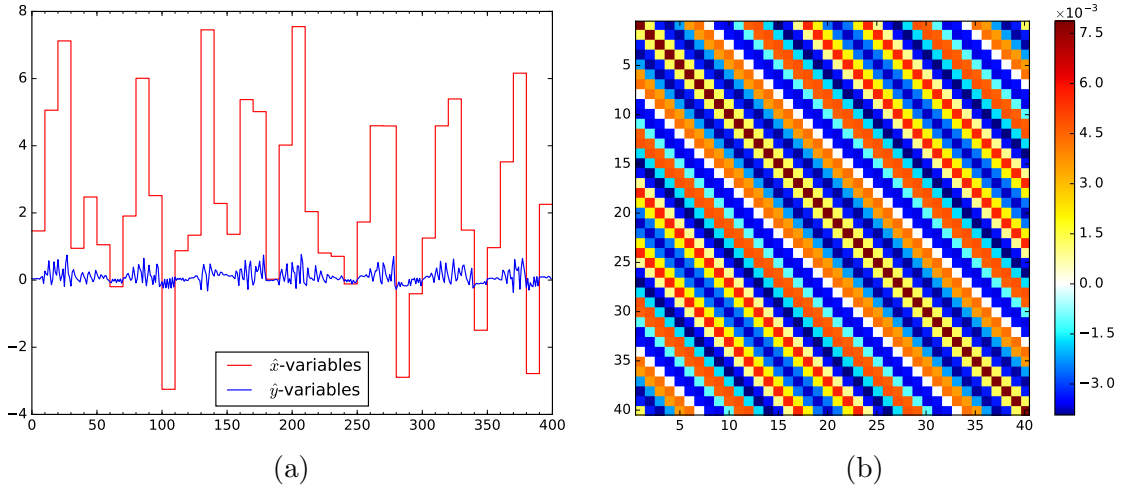


Figure 1.1: (a) Typical behavior of the LZ96 model, (b) The estimate of the true model error covariance matrix

as LZ96. The “true” state of the dynamical system is represented by the integration of (1.25) by the fourth-order Runge-Kutta method with  $b = c = 10$ ,  $h = 1$ ,  $K = 40$ ,  $J = 10$  and  $F = 8$ . By ignoring the effects of the  $\hat{y}_{jk}$  variables, the Lorenz 40-variable model [64]

$$\frac{dx_k}{dt} = x_{k-1}(x_{k+1} - x_{k-2}) - x_k + F \quad (1.26)$$

will only approximate the true state evolution and model error is now introduced by the unrepresented small-scale dynamics. Thus, for the data assimilation process, the true state  $\mathbf{x}_i^t$  at time  $t_i$  will be the  $\hat{x}$ -values produced from the integration of (1.25), whereas the forecast model  $\mathcal{M}_i$  will be the integration of (1.26) using a constant step-size  $\Delta t = 0.05$ , which identifies to a 6-hour time period. The integration of (1.25) requires a smaller time-step to preserve numerical stability, so a 6-hour forecast is achieved through ten smaller time-steps with  $\Delta t = 0.005$ .

For comparison purposes only, an estimate of the true model error covariance matrix  $\mathbf{Q}^t$  is obtained in the following way. First, a trajectory of 800,000 6-hour



forecasts using (1.25) is formed. The state where  $\hat{x}_k = 8$  for all  $k \neq K/2$ ,  $\hat{x}_{K/2} = 8.008$ , and  $\hat{y}_{jk} = 0$  is taken as the initial condition for the LZ96 model. Then, the  $x$ -values from each state is forecast using (1.26). The difference between  $\hat{x}$ - and  $x$ -values represents the error in the forecast model. The sample covariance of these differences is formed and represents the true time-invariant model error covariance  $\mathbf{Q}^t$ . Figure 1.1 shows the typical behavior of the LZ96 model and the estimated model error covariance matrix when the small-scale dynamics are ignored.

A demonstration of the evaluation of the tangent linear and adjoint models is done for the Lorenz 40-variable forecast function  $\mathcal{M}_i$ , the integration of (1.26) using the fourth-order Runge-Kutta method. For an autonomous differential equation system,

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}), \quad \mathbf{x}(t_0) = \mathbf{x}_0, \quad (1.27)$$

such as (1.26), the Runge-Kutta method numerically approximates the solution according to

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \frac{\Delta t}{6} [\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4], \quad (1.28)$$

where

$$\mathbf{k}_1 = \mathbf{f}(\mathbf{x}_i) \quad (1.29a)$$

$$\mathbf{k}_2 = \mathbf{f}\left(\mathbf{x}_i + \frac{\Delta t}{2}\mathbf{k}_1\right) \quad (1.29b)$$

$$\mathbf{k}_3 = \mathbf{f}\left(\mathbf{x}_i + \frac{\Delta t}{2}\mathbf{k}_2\right) \quad (1.29c)$$

$$\mathbf{k}_4 = \mathbf{f}(\mathbf{x}_i + \Delta t\mathbf{k}_3). \quad (1.29d)$$

The tangent linear and adjoint of (1.28) requires the tangent linear and adjoint of

the right-hand-side of (1.26). Starting with this,  $dx_k/dt$  will be denoted  $y_k$  for the ease of notation. The tangent linear model requires the derivatives of  $y_k$  with respect to all input variables. Since  $y_k$  depends on  $x_{k-2}$ ,  $x_{k-1}$ ,  $x_k$ , and  $x_{k+1}$ , only derivatives with respect to these four inputs is required.

$$\frac{\partial y_k}{\partial x_{k-2}} = -x_{k-1}, \quad \frac{\partial y_k}{\partial x_{k-1}} = x_{k+1} - x_{k-2}, \quad \frac{\partial y_k}{\partial x_k} = -1, \quad \frac{\partial y_k}{\partial x_{k+1}} = x_{k-1} \quad (1.30)$$

It follows from (1.23) that the tangent linear model of the right-hand-side of (1.26) is

$$\dot{y}_k = -x_{k-1}\dot{x}_{k-2} + (x_{k+1} - x_{k-2})\dot{x}_{k-1} - \dot{x}_k + x_{k-1}\dot{x}_{k+1}. \quad (1.31)$$

The tangent linear model can be rewritten to be more computationally efficient, reducing the number of operations required to compute the output.

$$\dot{y}_k = x_{k-1}(\dot{x}_{k+1} - \dot{x}_{k-2}) + \dot{x}_{k-1}(x_{k+1} - x_{k-2}) - \dot{x}_k \quad (1.32)$$

For the adjoint, the derivatives of all outputs which depend on  $x_k$  are needed. The outputs which depend on  $x_k$  are  $y_{k-1}$ ,  $y_k$ ,  $y_{k+1}$ , and  $y_{k+2}$ .

$$\frac{\partial y_{k-1}}{\partial x_k} = x_{k-2}, \quad \frac{\partial y_k}{\partial x_k} = -1, \quad \frac{\partial y_{k+1}}{\partial x_k} = x_{k+2} - x_{k-1}, \quad \frac{\partial y_{k+2}}{\partial x_k} = -x_{k+1} \quad (1.33)$$

Then, the adjoint model of the right-hand-side of (1.26) follows from (1.24).

$$\bar{x}_k = x_{k-2}\bar{y}_{k-1} - \bar{y}_k + (x_{k+2} - x_{k-1})\bar{y}_{k+1} - x_{k+1}\bar{y}_{k+2} \quad (1.34)$$

For the Runge-Kutta method (1.28), the tangent linear model is given by

$$\dot{\mathbf{x}}_{i+1} = \dot{\mathbf{x}}_i + \frac{\Delta t}{6} \left[ \dot{\mathbf{k}}_1 + 2\dot{\mathbf{k}}_2 + 2\dot{\mathbf{k}}_3 + \dot{\mathbf{k}}_4 \right], \quad (1.35)$$

where the  $\dot{\mathbf{k}}$  terms are obtained from the application of the chain rule

$$\dot{\mathbf{k}}_1 = \mathbf{F}(\mathbf{x}_i) \dot{\mathbf{x}}_i \quad (1.36a)$$

$$\dot{\mathbf{k}}_2 = \mathbf{F} \left( \mathbf{x}_i + \frac{\Delta t}{2} \mathbf{k}_1 \right) \left( \dot{\mathbf{x}}_i + \frac{\Delta t}{2} \dot{\mathbf{k}}_1 \right) \quad (1.36b)$$

$$\dot{\mathbf{k}}_3 = \mathbf{F} \left( \mathbf{x}_i + \frac{\Delta t}{2} \mathbf{k}_2 \right) \left( \dot{\mathbf{x}}_i + \frac{\Delta t}{2} \dot{\mathbf{k}}_2 \right) \quad (1.36c)$$

$$\dot{\mathbf{k}}_4 = \mathbf{F}(\mathbf{x}_i + \Delta t \mathbf{k}_3) \left( \dot{\mathbf{x}}_i + \Delta t \dot{\mathbf{k}}_3 \right) \quad (1.36d)$$

and  $\mathbf{F}$  denotes the tangent linear model of  $\mathbf{f}$ , that is (1.32) for the Lorenz model.

The adjoint of (1.28) is found by backward propagation the derivative of  $\mathbf{x}_{i+1}$  with respect to  $\mathbf{k}_4$ ,  $\mathbf{k}_3$ ,  $\mathbf{k}_2$ , and  $\mathbf{k}_1$ . By defining

$$\widehat{\mathbf{k}}_4 = \mathbf{F}^T(\mathbf{x}_i + \Delta t \mathbf{k}_3) \bar{\mathbf{x}}_{i+1} \quad (1.37a)$$

$$\widehat{\mathbf{k}}_3 = \mathbf{F}^T \left( \mathbf{x}_i + \frac{\Delta t}{2} \mathbf{k}_2 \right) \left( \bar{\mathbf{x}}_{i+1} + \frac{\Delta t}{2} \widehat{\mathbf{k}}_4 \right) \quad (1.37b)$$

$$\widehat{\mathbf{k}}_2 = \mathbf{F}^T \left( \mathbf{x}_i + \frac{\Delta t}{2} \mathbf{k}_1 \right) \left( \bar{\mathbf{x}}_{i+1} + \frac{\Delta t}{2} \widehat{\mathbf{k}}_3 \right) \quad (1.37c)$$

$$\widehat{\mathbf{k}}_1 = \mathbf{F}^T(\mathbf{x}_i) \left( \bar{\mathbf{x}}_{i+1} + \Delta t \widehat{\mathbf{k}}_2 \right), \quad (1.37d)$$

where  $\mathbf{F}^T$  denotes the adjoint of  $\mathbf{f}$ , which is (1.34) for the Lorenz model, the adjoint of (1.28) can be written as follows.

$$\bar{\mathbf{x}}_i = \bar{\mathbf{x}}_{i+1} + \frac{\Delta t}{6} \left[ \widehat{\mathbf{k}}_1 + 2\widehat{\mathbf{k}}_2 + 2\widehat{\mathbf{k}}_3 + \widehat{\mathbf{k}}_4 \right] \quad (1.38)$$

## 1.5 Organization

Chapter 2 provides an overview of w4D-Var, including the derivation of the cost functional and the analysis equations. Chapter 3 describes the Monte Carlo approach for estimating the model error statistics, the basis of which has been published in the journal *Procedia Computer Science* [72]. Chapter 4 provides research on statistical consistency diagnostics, extending previously published results. Chapter 5 shows the derivation of data assimilation sensitivity equations to a forecast error measure that are utilized in a gradient descent algorithm to reduce forecast error. The second stage of numerical experiments are conducted with discretized shallow water equations, more computationally demanding than the Lorenz model, in Chapter 6. Appendix A gives some notes on the computational implementation of the ensemble and hybrid error covariances, Appendix B shows the derivation of the discrete shallow water model using finite differences, and Appendix C gives the tangent linear and adjoint models for the discrete shallow water model.

## Chapter 2

### Weak-Constraint 4D-Var

Weak-constraint 4D-Var provides a sequence of time-distributed analyses  $\mathbf{x}_i^a \in \mathbb{R}^n$  that estimate the true state  $\mathbf{x}_i^t$  of a dynamical system at time  $t_i$  of the data assimilation interval  $[t_0, t_N]$  by solving a large-scale optimization problem. This involves finding the states that minimize the w4D-Var cost functional  $J(\mathbf{x}_0, \dots, \mathbf{x}_N)$ . The derivation of the cost functional is shown next.

#### 2.1 The Cost Functional

The w4D-Var optimization problem relies on a number of statistical assumptions [36, 59].

**Background Information:** Suppose that prior knowledge of the true state  $\mathbf{x}_0^t \in \mathbb{R}^n$  is known in the form of an unbiased estimate  $\mathbf{x}_0^b$ . Let

$$\boldsymbol{\varepsilon}^b = \mathbf{x}_0^t - \mathbf{x}_0^b \tag{2.1}$$

be the error in the estimate. Assume that  $\mathbf{x}_0^t$  has the multivariate normal distribution with mean  $\mathbf{x}_0^b$  and positive definite covariance matrix  $\mathbf{B}$ . That is,  $\mathbf{x}_0^t | \mathbf{x}_0^b \sim N(\mathbf{x}_0^b, \mathbf{B})$ ,

with probability density

$$p(\mathbf{x}_0|\mathbf{x}_0^b) \propto e^{-J_b(\mathbf{x}_0)} \quad (2.2)$$

where

$$J_b(\mathbf{x}_0) = \frac{1}{2}[\mathbf{x}_0 - \mathbf{x}_0^b]^\text{T} \mathbf{B}^{-1}[\mathbf{x}_0 - \mathbf{x}_0^b]. \quad (2.3)$$

**Observations:** Suppose that  $\mathbf{y}_i \in \mathbb{R}^{p_i}$  represents an observation about the true state  $\mathbf{x}_i^t$  at time  $t_i$ , for  $i = 0, 1, \dots, N$ . Let

$$\mathbf{y}_i = \mathbf{h}_i(\mathbf{x}_i^t) + \boldsymbol{\varepsilon}_i^o \quad (2.4)$$

where  $\mathbf{h}_i : \mathbb{R}^n \rightarrow \mathbb{R}^{p_i}$  is a vector-valued function and  $\boldsymbol{\varepsilon}_i^o$  is measurement noise, uncorrelated in time. Assume that  $\boldsymbol{\varepsilon}_i^o \sim N(\mathbf{0}, \mathbf{R}_i)$  with  $\mathbf{R}_i$  a positive definite matrix. Then  $\mathbf{y}_i|\mathbf{x}_i^t \sim N(\mathbf{h}_i(\mathbf{x}_i^t), \mathbf{R}_i)$ , with probability density

$$p(\mathbf{y}_i|\mathbf{x}_i) \propto e^{-J_{o,i}(\mathbf{x}_i)} \quad (2.5)$$

where

$$J_{o,i}(\mathbf{x}_i) = \frac{1}{2}[\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)]^\text{T} \mathbf{R}_i^{-1}[\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)]. \quad (2.6)$$

Then  $p(\mathbf{y}_0|\mathbf{x}_0) \cdots p(\mathbf{y}_N|\mathbf{x}_N) = J_o(\mathbf{x}_0, \dots, \mathbf{x}_N)$ , where  $J_o(\mathbf{x}_0, \dots, \mathbf{x}_N)$  is the sum of all the  $J_{o,i}(\mathbf{x}_i)$  so that

$$J_o(\mathbf{x}_0, \dots, \mathbf{x}_N) = \frac{1}{2} \sum_{i=0}^N [\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)]^\text{T} \mathbf{R}_i^{-1}[\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)]. \quad (2.7)$$

**Model:** Suppose that the unknown states  $\mathbf{x}_i^t \in \mathbb{R}^n$  are related by a model

$$\mathbf{x}_i^t = \mathcal{M}_i(\mathbf{x}_{i-1}^t) + \boldsymbol{\varepsilon}_i^q \quad (2.8)$$

where  $\boldsymbol{\varepsilon}_i^q \sim N(\mathbf{q}_i, \mathbf{Q}_i)$  is the error in the forecast model  $\mathcal{M}_i$ . Model error is uncorrelated in time and the vector  $\mathbf{q}_i$  is the model error bias, representing the systematic errors in  $\mathcal{M}_i$ .

The w4D-Var analysis is the sequence of states  $\{\mathbf{x}_0^a, \dots, \mathbf{x}_N^a\}$  that maximize the posterior distribution  $p(\mathbf{x}_0, \dots, \mathbf{x}_N | \mathbf{y}_0, \dots, \mathbf{y}_N)$ . Using Bayes' theorem,

$$p(\mathbf{x}_0, \dots, \mathbf{x}_N | \mathbf{y}_0, \dots, \mathbf{y}_N) = \frac{p(\mathbf{y}_0, \dots, \mathbf{y}_N | \mathbf{x}_0, \dots, \mathbf{x}_N) p(\mathbf{x}_0, \dots, \mathbf{x}_N)}{p(\mathbf{y}_0, \dots, \mathbf{y}_N)}. \quad (2.9)$$

Since the denominator term is independent of the  $\mathbf{x}_i$  and acts as a normalizing constant for  $p(\mathbf{x}_0, \dots, \mathbf{x}_N | \mathbf{y}_0, \dots, \mathbf{y}_N)$ , let  $p(\mathbf{y}_0, \dots, \mathbf{y}_N) = C^{-1}$ . Further, assume that the observations are independent given that one knows the true state [50, 80], so

$$p(\mathbf{y}_0, \dots, \mathbf{y}_N | \mathbf{x}_0, \dots, \mathbf{x}_N) = \prod_{i=0}^N p(\mathbf{y}_i | \mathbf{x}_i). \quad (2.10)$$

Also, assume that the sequence  $\{\mathbf{x}_0, \dots, \mathbf{x}_N\}$  is Markov, so that

$$p(\mathbf{x}_0, \dots, \mathbf{x}_N) = p(\mathbf{x}_0) \prod_{i=1}^N p(\mathbf{x}_i | \mathbf{x}_{i-1}). \quad (2.11)$$

By (2.8), it is noticed that  $\mathbf{x}_i^t | \mathbf{x}_{i-1}^t \sim N(\mathcal{M}_i(\mathbf{x}_{i-1}^t) + \mathbf{q}_i, \mathbf{Q}_i)$ . Thus  $p(\mathbf{x}_0, \dots, \mathbf{x}_N)$  is the product of normal distributions, with probability density

$$p(\mathbf{x}_0, \dots, \mathbf{x}_N) \propto e^{-J_b(\mathbf{x}_0)} e^{-J_q(\mathbf{x}_0, \dots, \mathbf{x}_N)}, \quad (2.12)$$

where

$$J_q(\mathbf{x}_0, \dots, \mathbf{x}_N) = \frac{1}{2} \sum_{i=1}^N [\mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1}) - \mathbf{q}_i]^T \mathbf{Q}_i^{-1} [\mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1}) - \mathbf{q}_i]. \quad (2.13)$$

Putting this all together,

$$p(\mathbf{x}_0, \dots, \mathbf{x}_N | \mathbf{y}_0, \dots, \mathbf{y}_N) = C' e^{-J(\mathbf{x}_0, \dots, \mathbf{x}_N)}, \quad (2.14)$$

where  $C'$  is a constant of proportionality and  $J(\mathbf{x}_0, \dots, \mathbf{x}_N)$  represents the w4D-Var cost functional (1.19). The analysis sequence  $\{\mathbf{x}_0^a, \dots, \mathbf{x}_N^a\}$  is found by minimizing (1.19) with respect to  $\{\mathbf{x}_0, \dots, \mathbf{x}_N\}$ .

Often in practice, model error is assumed to be unbiased, i.e.  $\mathbf{q}_i$  is set to zero in (1.19). By taking into account model error in w4D-Var, the control variable is the time-distributed sequence of states  $\{\mathbf{x}_0, \dots, \mathbf{x}_N\}$ . Trémolet [78] describes other possible formulations of the control variable, such as  $\{\mathbf{x}_0, \boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_N\}$ , where

$$\boldsymbol{\eta}_i = \mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1}) \quad (2.15)$$

represents the error in the forecast model  $\mathcal{M}_i$  that advances the state from time  $t_{i-1}$  to time  $t_i$ .

## 2.2 The Incremental Algorithm

Due to the nonlinearity of the observation operator  $\mathbf{h}_i$  and the discrete model  $\mathcal{M}_i$ , the incremental algorithm, introduced by Courtier et al. [13] for strong-constraint 4D-Var, can be adapted to minimize (1.19) through a sequence of quadratic approximations in which  $\mathbf{h}_i$  and  $\mathcal{M}_i$  are linearized about prior state estimates  $\mathbf{x}_i^g$  for  $i = 0, 1, \dots, N$ .



The observation operator is linearized at each time

$$\mathbf{h}_i(\mathbf{x}_i) \approx \mathbf{h}_i(\mathbf{x}_i^g) + \mathbf{H}_i^g(\mathbf{x}_i - \mathbf{x}_i^g) \quad (2.16)$$

where  $\mathbf{H}_i^g$  denotes the  $p_i \times n$  Jacobian matrix of  $\mathbf{h}_i$  evaluated at  $\mathbf{x}_i^g$ . Similarly, for  $i = 1, \dots, N$ , linearization of the model forecast operator from time  $t_{i-1}$  to time  $t_i$  gives

$$\mathcal{M}_i(\mathbf{x}_{i-1}) \approx \mathcal{M}_i(\mathbf{x}_{i-1}^g) + \mathbf{M}_i^g(\mathbf{x}_{i-1} - \mathbf{x}_{i-1}^g), \quad (2.17)$$

where  $\mathbf{M}_i^g$  denotes the  $n \times n$  Jacobian matrix of  $\mathcal{M}_i$  evaluated at  $\mathbf{x}_{i-1}^g$ . From (2.16) and (2.17), the quadratic approximation to (1.19) can be written in terms of the increments  $\delta \mathbf{x}_i = \mathbf{x}_i - \mathbf{x}_i^g$ . By defining  $\delta \mathbf{x}_0^b = \mathbf{x}_0^b - \mathbf{x}_0^g$ ,  $\mathbf{d}_i = \mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i^g)$ , and  $\boldsymbol{\eta}_i^g = \mathbf{x}_i^g - \mathcal{M}_i(\mathbf{x}_{i-1}^g)$ , it is noticed that

$$\mathbf{x}_0 - \mathbf{x}_0^b = \delta \mathbf{x}_0 - \delta \mathbf{x}_0^b \quad (2.18)$$

$$\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i) \approx \mathbf{d}_i - \mathbf{H}_i^g \delta \mathbf{x}_i \quad (2.19)$$

$$\mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1}) \approx \delta \mathbf{x}_i - \mathbf{M}_i^g \delta \mathbf{x}_{i-1} + \boldsymbol{\eta}_i^g. \quad (2.20)$$

Equations (2.18)–(2.20) are used to approximate the nonlinear functional (1.19) by a quadratic functional that is minimized with respect to  $\{\delta \mathbf{x}_0, \dots, \delta \mathbf{x}_N\}$ .

$$\begin{aligned} \hat{J}(\delta \mathbf{x}_0, \dots, \delta \mathbf{x}_N) &= \frac{1}{2} [\delta \mathbf{x}_0 - \delta \mathbf{x}_0^b]^T \mathbf{B}^{-1} [\delta \mathbf{x}_0 - \delta \mathbf{x}_0^b] \\ &+ \frac{1}{2} \sum_{i=0}^N [\mathbf{d}_i - \mathbf{H}_i^g \delta \mathbf{x}_i]^T \mathbf{R}_i^{-1} [\mathbf{d}_i - \mathbf{H}_i^g \delta \mathbf{x}_i] \\ &+ \frac{1}{2} \sum_{i=1}^N [\delta \mathbf{x}_i - \mathbf{M}_i^g \delta \mathbf{x}_{i-1} + \boldsymbol{\eta}_i^g - \mathbf{q}_i]^T \mathbf{Q}_i^{-1} [\delta \mathbf{x}_i - \mathbf{M}_i^g \delta \mathbf{x}_{i-1} + \boldsymbol{\eta}_i^g - \mathbf{q}_i] \end{aligned} \quad (2.21)$$

An equivalent formulation to the quadratic cost functional (2.21) may be given as

$$\begin{aligned}\hat{J}(\delta \mathbf{x}) &= \frac{1}{2} [(\mathbf{F}^g)^{-1} \delta \mathbf{x} - \mathbf{g} - \mathbf{q}]^T \mathbf{P}^{-1} [(\mathbf{F}^g)^{-1} \delta \mathbf{x} - \mathbf{g} - \mathbf{q}] \\ &\quad + \frac{1}{2} [\mathbf{d} - \mathbf{H}^g \delta \mathbf{x}]^T \mathbf{R}^{-1} [\mathbf{d} - \mathbf{H}^g \delta \mathbf{x}],\end{aligned}\tag{2.22}$$

by defining the four-dimensional increment vector

$$\delta \mathbf{x} = \begin{bmatrix} \delta \mathbf{x}_0^T & \cdots & \delta \mathbf{x}_N^T \end{bmatrix}^T,\tag{2.23}$$

the four-dimensional innovation vector

$$\mathbf{d} = \mathbf{y} - \mathbf{h}(\mathbf{x}^g) = \begin{bmatrix} \mathbf{d}_0^T & \cdots & \mathbf{d}_N^T \end{bmatrix}^T\tag{2.24}$$

of dimension  $p = p_0 + p_1 + \cdots + p_N$ , the time-distributed linearized observation operator

$$\mathbf{H}^g = \begin{bmatrix} \mathbf{H}_0^g & & \\ & \ddots & \\ & & \mathbf{H}_N^g \end{bmatrix},\tag{2.25}$$

the block observation error covariance matrix

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_0 & & \\ & \ddots & \\ & & \mathbf{R}_N \end{bmatrix},\tag{2.26}$$

the block-diagonal matrix

$$\mathbf{P} = \begin{bmatrix} \mathbf{B} & & & \\ & \mathbf{Q}_1 & & \\ & & \ddots & \\ & & & \mathbf{Q}_N \end{bmatrix}, \quad (2.27)$$

the guess vector

$$\mathbf{g} = \begin{bmatrix} (\delta \mathbf{x}_0^b)^T & -(\boldsymbol{\eta}_1^g)^T & \cdots & -(\boldsymbol{\eta}_N^g)^T \end{bmatrix}^T, \quad (2.28)$$

and the bias vector

$$\mathbf{q} = \begin{bmatrix} \mathbf{0}^T & \mathbf{q}_1^T & \cdots & \mathbf{q}_N^T \end{bmatrix}^T. \quad (2.29)$$

The matrix

$$\mathbf{F}^g = \begin{bmatrix} \mathbf{I} & & & & \\ \mathbf{M}_1^g & \mathbf{I} & & & \\ \mathbf{M}_{0 \rightarrow 2}^g & \mathbf{M}_2^g & \mathbf{I} & & \\ \vdots & \vdots & \ddots & \ddots & \\ \mathbf{M}_{0 \rightarrow N}^g & \mathbf{M}_{1 \rightarrow N}^g & \cdots & \mathbf{M}_N^g & \mathbf{I} \end{bmatrix}, \quad (2.30)$$

where

$$\mathbf{M}_{i \rightarrow i+k}^g = \mathbf{M}_{i+k}^g \cdots \mathbf{M}_{i+2}^g \mathbf{M}_{i+1}^g \quad (2.31)$$

represents the tangent linear model integrated from time  $t_i$  to  $t_{i+k}$  evaluated at the guess states  $\mathbf{x}_i^g$ , comes from combining the background error and model error terms

of (2.21) into the block formulation of (2.22). It is noticed that  $\mathbf{F}^g$  is invertible with inverse

$$(\mathbf{F}^g)^{-1} = \begin{bmatrix} \mathbf{I} & & & & \\ -\mathbf{M}_1^g & \mathbf{I} & & & \\ & -\mathbf{M}_2^g & \mathbf{I} & & \\ & & \ddots & \ddots & \\ & & & -\mathbf{M}_N^g & \mathbf{I} \end{bmatrix}, \quad (2.32)$$

The analysis sequence is obtained by the minimization of the quadratic approximation (2.22) to the nonlinear cost functional (1.19) with respect to the four-dimensional increment  $\delta \mathbf{x}$ .

### 2.3 The Analysis Equation

To find the optimum increment which minimizes the cost functional (2.22), differentiate with respect to  $\delta \mathbf{x}$

$$\nabla_{\delta \mathbf{x}} \hat{J}(\delta \mathbf{x}) = (\mathbf{F}^g)^{-T} \mathbf{P}^{-1} [(\mathbf{F}^g)^{-1} \delta \mathbf{x} - \mathbf{g} - \mathbf{q}] - (\mathbf{H}^g)^T \mathbf{R}^{-1} [\mathbf{d} - \mathbf{H}^g \delta \mathbf{x}], \quad (2.33)$$

set the result equal to zero, and solve for  $\delta \mathbf{x}$ . Thus, the optimal increment  $\delta \mathbf{x}^a$  minimizing (2.22) solves the linear system

$$[(\mathbf{F}^g)^{-T} \mathbf{P}^{-1} (\mathbf{F}^g)^{-1} + (\mathbf{H}^g)^T \mathbf{R}^{-1} \mathbf{H}^g] \delta \mathbf{x}^a = (\mathbf{F}^g)^{-T} \mathbf{P}^{-1} [\mathbf{g} + \mathbf{q}] + (\mathbf{H}^g)^T \mathbf{R}^{-1} \mathbf{d}. \quad (2.34)$$

The theoretical solution to this problem is

$$\begin{aligned} \delta \mathbf{x}^a = & \left[ (\mathbf{F}^g)^{-T} \mathbf{P}^{-1} (\mathbf{F}^g)^{-1} + (\mathbf{H}^g)^T \mathbf{R}^{-1} \mathbf{H}^g \right]^{-1} \\ & \times \left[ (\mathbf{F}^g)^{-T} \mathbf{P}^{-1} [\mathbf{g} + \mathbf{q}] + (\mathbf{H}^g)^T \mathbf{R}^{-1} \mathbf{d} \right], \end{aligned} \quad (2.35)$$

in which the matrix inversion can be done by using the Sherman-Morrison-Woodbury matrix identity

$$[\mathbf{A} + \mathbf{C} \mathbf{B} \mathbf{D}^T]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{C} [\mathbf{B}^{-1} + \mathbf{D}^T \mathbf{A}^{-1} \mathbf{C}]^{-1} \mathbf{D}^T \mathbf{A}^{-1}, \quad (2.36)$$

where the matrices  $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$  have appropriate dimensions to make the expression valid. To use (2.36), identify  $(\mathbf{F}^g)^{-T} \mathbf{P}^{-1} (\mathbf{F}^g)^{-1}$  with  $\mathbf{A}$ ,  $(\mathbf{H}^g)^T$  with  $\mathbf{C}$ ,  $\mathbf{R}^{-1}$  with  $\mathbf{B}$ , and  $(\mathbf{H}^g)^T$  with  $\mathbf{D}$ . The optimal analysis increment  $\delta \mathbf{x}^a$  minimizes (2.21) and may be expressed as

$$\delta \mathbf{x}^a = \mathbf{K} [\mathbf{y} - \mathbf{h}(\mathbf{x}^g)] + [\mathbf{I} - \mathbf{K} \mathbf{H}^g] \mathbf{F}^g [\mathbf{g} + \mathbf{q}] \quad (2.37)$$

where the gain matrix  $\mathbf{K}$  is defined as

$$\mathbf{K} = \mathbf{F}^g \mathbf{P} (\mathbf{F}^g)^T (\mathbf{H}^g)^T [\mathbf{H}^g \mathbf{F}^g \mathbf{P} (\mathbf{F}^g)^T (\mathbf{H}^g)^T + \mathbf{R}]^{-1} \quad (2.38)$$

and  $\mathbf{h}$  is the four-dimensional observation operator

$$\mathbf{h}(\mathbf{x}) = \begin{bmatrix} \mathbf{h}_0^T(\mathbf{x}_0) & \cdots & \mathbf{h}_N^T(\mathbf{x}_N) \end{bmatrix}^T. \quad (2.39)$$

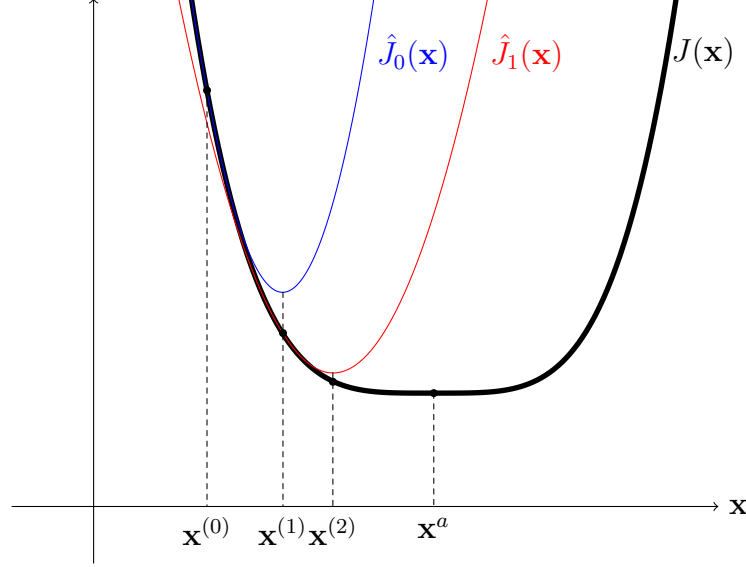


Figure 2.1: A demonstration of the incremental method showing the successive quadratic approximations of an example cost function.

Therefore, the four-dimensional state which minimizes the quadratic cost functional (2.21) is

$$\mathbf{x}^a = \mathbf{x}^g + \mathbf{K}[\mathbf{y} - \mathbf{h}(\mathbf{x}^g)] + [\mathbf{I} - \mathbf{K}\mathbf{H}^g]\mathbf{F}^g[\mathbf{g} + \mathbf{q}]. \quad (2.40)$$

For a single outer loop iteration, equation (2.40) gives the analysis state for the data assimilation window.

Once the analysis estimate is obtained, the incremental method can be done again with  $\mathbf{x}^a$  as the new guess. See Figure 2.1 for a visual of the successive quadratic approximations of the cost function. If this process is to be done again with  $\mathbf{x}^a$  as the new  $\mathbf{x}^g$ , the linearizations will need to be redone as to be centered at the new  $\mathbf{x}^g$ . Continue this process until some desired accuracy or a maximum number of iterations has been reached. Typically, it is impractical to do more than a few iterations of the incremental algorithm. A good choice for  $\mathbf{x}_i^g$  is  $\mathbf{x}_0^g = \mathbf{x}_0^b$  and  $\mathbf{x}_i^g = \mathcal{M}_i(\mathbf{x}_{i-1}^g) + \mathbf{q}_i$  since

this choice of guesses makes  $\mathbf{g} + \mathbf{q} = \mathbf{0}$ , in which case (2.40) becomes

$$\mathbf{x}^a = \mathbf{x}^g + \mathbf{K}[\mathbf{y} - \mathbf{h}(\mathbf{x}^g)] \quad (2.41)$$

which resembles the form of the Kalman filter.

It is usually desirable to find the analysis (2.40) by introducing a change of variables, such as

$$\boldsymbol{\chi} = \mathbf{P}^{-1/2}[(\mathbf{F}^g)^{-1}\delta\mathbf{x} - \mathbf{g} - \mathbf{q}], \quad (2.42)$$

for the purpose of preconditioning. With this variable change, (2.22) becomes a function of  $\boldsymbol{\chi}$

$$\begin{aligned} \hat{J}(\boldsymbol{\chi}) &= \frac{1}{2}\boldsymbol{\chi}^T\boldsymbol{\chi} \\ &+ \frac{1}{2}[\mathbf{H}^g\mathbf{F}^g(\mathbf{P}^{1/2}\boldsymbol{\chi} + \mathbf{g} + \mathbf{q}) - \mathbf{d}]^T\mathbf{R}^{-1}[\mathbf{H}^g\mathbf{F}^g(\mathbf{P}^{1/2}\boldsymbol{\chi} + \mathbf{g} + \mathbf{q}) - \mathbf{d}] \end{aligned} \quad (2.43)$$

and the minimization is performed to obtain  $\boldsymbol{\chi}^a$ , from which it is seen that

$$\delta\mathbf{x}^a = \mathbf{F}^g(\mathbf{P}^{1/2}\boldsymbol{\chi}^a + \mathbf{g} + \mathbf{q}). \quad (2.44)$$

The gradient of the cost  $\hat{J}$  with respect to  $\boldsymbol{\chi}$  is

$$\nabla_{\boldsymbol{\chi}}\hat{J}(\boldsymbol{\chi}) = \boldsymbol{\chi} + \mathbf{P}^{T/2}(\mathbf{F}^g)^T(\mathbf{H}^g)^T\mathbf{R}^{-1}[\mathbf{H}^g\mathbf{F}^g(\mathbf{P}^{1/2}\boldsymbol{\chi} + \mathbf{g} + \mathbf{q}) - \mathbf{d}]. \quad (2.45)$$

Setting (2.45) equal to zero, the linear system to solve is now

$$\begin{aligned} [\mathbf{I} + \mathbf{P}^{T/2}(\mathbf{F}^g)^T(\mathbf{H}^g)^T\mathbf{R}^{-1}\mathbf{H}^g\mathbf{F}^g\mathbf{P}^{1/2}]\boldsymbol{\chi}^a = \\ \mathbf{P}^{T/2}(\mathbf{F}^g)^T(\mathbf{H}^g)^T\mathbf{R}^{-1}[\mathbf{d} - \mathbf{H}^g\mathbf{F}^g(\mathbf{g} + \mathbf{q})] \end{aligned} \quad (2.46)$$

with solution

$$\boldsymbol{\chi}^a = \mathbf{P}^{T/2}(\mathbf{F}^g)^T(\mathbf{H}^g)^T[\mathbf{H}^g\mathbf{F}^g\mathbf{P}(\mathbf{F}^g)^T(\mathbf{H}^g)^T + \mathbf{R}]^{-1}[\mathbf{d} - \mathbf{H}^g\mathbf{F}^g(\mathbf{g} + \mathbf{q})]. \quad (2.47)$$

Since the matrix  $\mathbf{P}^{T/2}(\mathbf{F}^g)^T(\mathbf{H}^g)^T\mathbf{R}^{-1}\mathbf{H}^g\mathbf{F}^g\mathbf{P}^{1/2}$  is positive semidefinite, its eigenvalues are all nonnegative. Thus the eigenvalues of  $\mathbf{I} + \mathbf{P}^{T/2}(\mathbf{F}^g)^T(\mathbf{H}^g)^T\mathbf{R}^{-1}\mathbf{H}^g\mathbf{F}^g\mathbf{P}^{1/2}$  are all greater than or equal to 1, and so the matrix is positive definite, whereas  $(\mathbf{F}^g)^{-T}\mathbf{P}^{-1}(\mathbf{F}^g)^{-1} + (\mathbf{H}^g)^T\mathbf{R}^{-1}\mathbf{H}^g$  may not be positive definite. This is where the advantage of preconditioning with  $\boldsymbol{\chi}$  becomes apparent.

The analysis  $\mathbf{x}^a$  can also be obtained from the observation space evaluation of (2.40) via the two-step process of solving the linear system

$$[\mathbf{H}^g\mathbf{F}^g\mathbf{P}(\mathbf{F}^g)^T(\mathbf{H}^g)^T + \mathbf{R}]\mathbf{z} = \mathbf{d} - \mathbf{H}^g\mathbf{F}^g[\mathbf{g} + \mathbf{q}] \quad (2.48)$$

for the vector  $\mathbf{z} \in \mathbb{R}^p$  followed by the multiplication in

$$\mathbf{x}^a = \mathbf{x}^g + \mathbf{F}^g\mathbf{q} + \mathbf{F}^g\mathbf{P}(\mathbf{F}^g)^T(\mathbf{H}^g)^T\mathbf{z}. \quad (2.49)$$

When the number of observations  $p$  is less than the number of state space variables  $(N+1)n$ , obtaining the analysis  $\mathbf{x}^a$  via  $\mathbf{z}$  can be less computationally expensive due to the reduction in the dimension of the problem. The observation space implementation of the analysis scheme, including the outer loop iteration, is presented in the work of



Rosmond and Xu [70].

## Chapter 3

### Ensemble Data Assimilation

Ensemble data assimilation has been used in conjunction with variational methods in an attempt to capture the “errors of the day” and dynamically update the background error covariance, since error covariances are typically not updated between 4D-Var assimilation cycles. Evensen [31] introduces this Monte Carlo alternative as the ensemble Kalman filter (EnKF) and it has since been implemented in various studies, e.g. [43, 48, 49]. Lorenc [61] and Fairbairn et al. [35] investigate the potential use of EnKF for NWP applications and its analysis performance, as compared with 4D-Var. Ensemble data assimilation can estimate not only the model error covariance matrices, but also bias. Traditionally, an assumption is made that the errors in data assimilation are unbiased to simplify the computational cost or because the information about error biases is not available. Bias in data assimilation has been explored in the works by Dee [21], Dee and Da Silva [22], and Dee and Todling [23], where it is noted that errors in models and the data are often systematic rather than random. Attempts to correct for error bias have been made in the form of bias detection and correction methods and “bias-aware” data assimilation methods, including bias correction in variational data assimilation [24], but not in the context of w4D-Var. Bias-aware Kalman filters have been explored by Drécourt et al. [29].

Ensembles can also be used to estimate the model error covariances  $\mathbf{Q}_i$  and the model biases  $\mathbf{q}_i$  by using ensembles for the analysis states  $\mathbf{x}_{i,j}^a$ . The steps needed to

obtain ensemble estimates of model error are presented next [72].

### 3.1 Derivation of the Model Error Ensemble

When the true model error statistics are unknown, the data assimilation system specifications of the error bias  $\mathbf{q}_i \neq \mathbf{q}_i^t$  and error covariance  $\mathbf{Q}_i \neq \mathbf{Q}_i^t$  are made, where the superscript  $()^t$  is used to distinguish the true error statistics  $\mathbf{Q}_i^t, \mathbf{q}_i^t$  from their specified counterparts  $\mathbf{Q}_i, \mathbf{q}_i$ . The incremental algorithm is used to perform w4D-Var over a window  $[t_0, t_N]$ . The states  $\mathbf{x}_i^g$  at which the model and observation operators are linearized will utilize the assumed model error statistics, that is, let

$$\mathbf{x}_0^g = \mathbf{x}_0^b \quad (3.1a)$$

$$\mathbf{x}_i^g = \mathcal{M}_i(\mathbf{x}_{i-1}^g) + \mathbf{q}_i, \quad i = 1, \dots, N. \quad (3.1b)$$

The incremental method then produces the four-dimensional analysis given by (2.41).

An ensemble of analyses  $\mathbf{x}_{i,j}^a$ , where  $i = 0, 1, \dots, N$  and  $j = 1, \dots, N_e$ , is used to produce a low-rank representation to the model error covariance. The setup is as follows.

- Prescribe the error statistics  $\mathbf{B}$ ,  $\mathbf{R}_i$ ,  $\mathbf{Q}_i$ , and  $\mathbf{q}_i$  to be used for each w4D-Var problem, the same specification for each ensemble member.
- From the background state  $\mathbf{x}_0^b$ , form the background ensemble  $\mathbf{x}_{0,j}^b = \mathbf{x}_0^b + \boldsymbol{\varepsilon}_j^b$ , where  $\boldsymbol{\varepsilon}_j^b$  is generated from the normal distribution  $N(\mathbf{0}, \mathbf{B})$ .
- Perturb the observation  $\mathbf{y}_i$  to form an ensemble  $\mathbf{y}_{i,j} = \mathbf{y}_i + \boldsymbol{\varepsilon}_{i,j}^o$ , where the perturbation  $\boldsymbol{\varepsilon}_{i,j}^o$  is normally distributed with mean zero and covariance  $\mathbf{R}_i$ , for  $i = 0, 1, \dots, N$  and  $j = 1, \dots, N_e$ .

- For each member of the background ensemble  $\mathbf{x}_{0,j}^b$ , form the corresponding ensemble of guesses  $\mathbf{x}_{i,j}^g$  according to (3.1) using the assumed model error bias  $\mathbf{q}_i$ .

A substitute for using the statistics of  $\mathbf{B}$  to perturb the background is to use the approximate background error

$$\varepsilon = \|\mathbf{x}_0^a - \mathbf{x}_0^b\|/n \quad (3.2)$$

as the standard deviation for the mean zero normally distributed perturbation. A multiplicative constant  $\beta$  can be included so that the standard deviation of the perturbations is  $\beta\varepsilon$ .

It is noted here that perturbing observations was introduced by Burgers et al. [8] and has been implemented in several experiments, such as work done by Pereira and Berre [69]. Whitaker and Hamill [79] show that ensemble data assimilation can be performed without perturbing the observations, however, in an ensemble square root filter.

By performing w4D-Var using the incremental method with data  $\mathbf{x}_{0,j}^g, \dots, \mathbf{x}_{N,j}^g$  and observations  $\mathbf{y}_{0,j}, \dots, \mathbf{y}_{N,j}$ , we get an ensemble of analysis states  $\mathbf{x}_{i,j}^a$ . The four-dimensional analysis ensemble  $\mathbf{x}_j^a$  follows from (2.41)

$$\mathbf{x}_j^a = \mathbf{x}_j^g + \mathbf{K}_j[\mathbf{y}_j - \mathbf{h}(\mathbf{x}_j^g)] \quad (3.3)$$

where the gain matrix  $\mathbf{K}_j$  may vary with the ensemble member  $j$ . In this framework, the best estimate of the true state is obtained as the ensemble average for each time

$$\bar{\mathbf{x}}_i^a = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathbf{x}_{i,j}^a. \quad (3.4)$$

From (3.3) and (3.4), define ensemble estimates to model error

$$\boldsymbol{\eta}_{i,j} = \bar{\mathbf{x}}_i^a - \mathcal{M}_i(\mathbf{x}_{i-1,j}^a) \quad (3.5)$$

for  $i = 1, \dots, N$  and  $j = 1, \dots, N_e$ . With the model error ensemble now available, the model error bias  $\mathbf{q}_i^t$  is estimated by the ensemble mean

$$\mathbf{q}_{i,e} = \frac{1}{N_e} \sum_{j=1}^{N_e} \boldsymbol{\eta}_{i,j} \quad (3.6)$$

and the associated ensemble estimates to the model error covariance matrix are

$$\mathbf{Q}_{i,e} = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} [\boldsymbol{\eta}_{i,j} - \mathbf{q}_{i,e}][\boldsymbol{\eta}_{i,j} - \mathbf{q}_{i,e}]^T. \quad (3.7)$$

Bickel and Levina [5] discuss regularized estimates of covariance matrices by banding the sample covariance matrix. Now that estimates for the model error bias and model error covariance are available,  $\mathbf{q}_{i,e}$  and  $\mathbf{Q}_{i,e}$  may be used in a w4D-Var data assimilation system. This procedure is summarized in Algorithm 3.1. Possible ensemble-based assimilation schemes are described next.

### 3.2 Ensemble-based w4D-Var Schemes

Instead of prescribing the model error covariance matrices as static  $\mathbf{Q}_i$  that do not change between assimilation cycles, one approach is to specify the model error bias and model error covariance matrices as  $\mathbf{q}_{i,e}$  and  $\mathbf{Q}_{i,e}$ , respectively. This choice of specifying  $\mathbf{q}_i = \mathbf{q}_{i,e}$  and  $\mathbf{Q}_i = \mathbf{Q}_{i,e}$  utilizes the information from the “errors of the day” to improve the quality of the analysis. These specifications can be kept up-to-date in future time-steps by computing the ensemble estimates  $\mathbf{q}_{i,e}$  and  $\mathbf{Q}_{i,e}$  in each

---

**Algorithm 3.1:** Computation of the ensemble estimates of model error.

---

```

1: procedure MODEL ERROR ENSEMBLE( $\mathbf{B}$ ,  $\mathbf{R}_i$ ,  $\mathbf{Q}_i$ ,  $\mathbf{q}_i$ ,  $N_e$ )
2:   for  $j = 1, \dots, N_e$  do
3:      $\mathbf{x}_{0,j}^b = \mathbf{x}_0^b + \boldsymbol{\varepsilon}_j^b$  ▷ Perturb the background
4:      $\mathbf{x}_{0,j}^g = \mathbf{x}_{0,j}^b$  ▷ Set the guess states

5:     for  $i = 1, \dots, N$  do
6:        $\mathbf{x}_{i,j}^g = \mathcal{M}_i(\mathbf{x}_{i-1,j}^g) + \mathbf{q}_i$ 
7:     end for

8:     for  $i = 0, 1, \dots, N$  do
9:        $\mathbf{y}_{i,j} = \mathbf{y}_i + \boldsymbol{\varepsilon}_{i,j}^o$  ▷ Perturb the observations
10:    end for

11:     $(\mathbf{x}_{0,j}^a, \dots, \mathbf{x}_{N,j}^a) = \text{w4DVar}(\mathbf{B}, \mathbf{R}_i, \mathbf{Q}_i, \mathbf{q}_i, \mathbf{x}_{i,j}^g, \mathbf{y}_{i,j})$  ▷ Analysis ensemble
12:  end for

13:  for  $i = 0, 1, \dots, N$  do
14:    Compute  $\bar{\mathbf{x}}_i^a$  from equation (3.4)

15:    for  $j = 1, \dots, N_e$  do
16:      Compute  $\boldsymbol{\eta}_{i,j}$  from equation (3.5)
17:    end for
18:  end for

19:  for  $i = 1, \dots, N$  do
20:    Compute  $\mathbf{q}_{i,e}$  from equation (3.6)
21:    Compute  $\mathbf{Q}_{i,e}$  from equation (3.7)
22:  end for
23: end procedure

```

---

data assimilation cycle.

The ensemble covariance matrices  $\mathbf{Q}_{i,e}$  may have low rank due to a small ensemble size and additionally suffer from the presence of sampling error. To reduce this, one may replace  $\mathbf{Q}_i$  in the data assimilation system by the Hadamard product of the ensemble covariance  $\mathbf{Q}_{i,e}$  with a localization matrix

$$\mathbf{Q}_i = \mathbf{Q}_{i,e} \circ \mathbf{C}_i \quad (3.8)$$

where  $\mathbf{C}_i$  is a properly selected correlation matrix. A popular correlation function to apply is the fifth-order rational function of compact support

$$C_0(z, 1/2, c) = \begin{cases} -\frac{1}{4} \left(\frac{|z|}{c}\right)^5 + \frac{1}{2} \left(\frac{z}{c}\right)^4 + \frac{5}{8} \left(\frac{|z|}{c}\right)^3 - \frac{5}{3} \left(\frac{z}{c}\right)^2 + 1, & 0 \leq |z| \leq c \\ \frac{1}{12} \left(\frac{|z|}{c}\right)^5 - \frac{1}{2} \left(\frac{z}{c}\right)^4 + \frac{5}{8} \left(\frac{|z|}{c}\right)^3 + \frac{5}{3} \left(\frac{z}{c}\right)^2 - \\ \quad 5\frac{|z|}{c} + 4 - \frac{2c}{3|z|}, & c < |z| \leq 2c, \\ 0, & |z| > 2c \end{cases} \quad (3.9)$$

given by equation (4.10) of Gaspari and Cohn [37].

Another option is to specify the model error covariance matrices as a linear combination of a static matrix  $\mathbf{Q}_{i,c}$  and the ensemble covariance

$$\mathbf{Q}_i = \alpha_i \mathbf{Q}_{i,c} + (1 - \alpha_i) \mathbf{Q}_{i,e}. \quad (3.10)$$

A localization matrix  $\mathbf{C}_i$  may applied to  $\mathbf{Q}_{i,e}$  so that (3.10) is replaced by

$$\mathbf{Q}_i = \alpha_i \mathbf{Q}_{i,c} + (1 - \alpha_i) \mathbf{Q}_{i,e} \circ \mathbf{C}_i. \quad (3.11)$$

Similarly, the model error bias is specified as a linear combination of a static vector  $\mathbf{q}_{i,c}$  and the ensemble average using the same parameter

$$\mathbf{q}_i = \alpha_i \mathbf{q}_{i,c} + (1 - \alpha_i) \mathbf{q}_{i,e} \quad (3.12)$$

where  $0 \leq \alpha_i \leq 1$ . This combination of two specifications of model error is referred to as hybrid data assimilation. The work by Efron and Morris [30] is at the origin of estimating covariance matrices and the hybrid approach. Estimating a covariance matrix using a linear combination of the sample covariance matrix and the identity

matrix was introduced by Ledoit and Wolf [57]. For  $\alpha_i = 1$ , the specified model error will utilize the current static specification, or the status quo, while for  $\alpha_i = 0$ , it will be set to the ensemble model error statistics. Hybrid data assimilation is designed to combine the merits of both the static component  $\{\mathbf{Q}_{i,c}, \mathbf{q}_{i,c}\}$  and the dynamic component  $\{\mathbf{Q}_{i,e}, \mathbf{q}_{i,e}\}$  with a value of  $\alpha_i$  satisfying  $0 < \alpha_i < 1$  to improve the quality of the analysis more than the static and dynamic components can do alone. Appendix A contains some notes on the implementation of the ensemble and hybrid error covariances.

### 3.3 Ensemble Error Statistics

This section provides an analysis of the statistical properties of the model error ensemble average  $\mathbf{q}_{i,e}$ . The average  $\mathbf{q}_{i,e}$  plays an important role in the ensemble-based schemes described in the previous section, so it seems that an exploration of its properties is in order.

To better understand how well  $\mathbf{q}_{i,e}$  estimates the model bias  $\mathbf{q}_i^t$ , it is necessary to investigate the expectation  $E[\boldsymbol{\varepsilon}_i^q - \mathbf{q}_{i,e}]$ . From the linearization of  $\mathcal{M}_i$  at  $\bar{\mathbf{x}}_{i-1}^a$  and by (3.5), it is noticed that (3.6) becomes

$$\mathbf{q}_{i,e} \approx \bar{\mathbf{x}}_i^a - \mathcal{M}_i(\bar{\mathbf{x}}_{i-1}^a). \quad (3.13)$$

To a first-order approximation, (3.13) implies

$$\boldsymbol{\varepsilon}_i^q - \mathbf{q}_{i,e} \approx (\mathbf{x}_i^t - \bar{\mathbf{x}}_i^a) - \mathbf{M}_i^a[\mathbf{x}_{i-1}^t - \bar{\mathbf{x}}_{i-1}^a] \quad (3.14)$$

where  $\mathbf{M}_i^a$  represents the linearized model evaluated at  $\bar{\mathbf{x}}_{i-1}^a$ .

Since (3.14) involves analysis errors at time  $t_i$  and  $t_{i-1}$ , it will be convenient to



analyze all of the differences  $\boldsymbol{\varepsilon}_i^q - \mathbf{q}_{i,e}$  simultaneously using the four-dimensional model error

$$\boldsymbol{\varepsilon}^q = \begin{bmatrix} (\boldsymbol{\varepsilon}_1^q)^T & (\boldsymbol{\varepsilon}_2^q)^T & \cdots & (\boldsymbol{\varepsilon}_N^q)^T \end{bmatrix}^T, \quad (3.15)$$

and the four-dimensional model error ensemble

$$\mathbf{q}_e = \begin{bmatrix} \mathbf{q}_{1,e}^T & \mathbf{q}_{2,e}^T & \cdots & \mathbf{q}_{N,e}^T \end{bmatrix}^T, \quad (3.16)$$

where it is noticed that the expectation  $E[\boldsymbol{\varepsilon}^q - \mathbf{q}_e]$  depends on  $E[\mathbf{x}_i^t - \bar{\mathbf{x}}_i^a]$  for  $i = 0, \dots, N$ . Since the perturbations introduced in the ensemble data assimilation are unbiased, it can be shown that  $E[\mathbf{x}_i^t - \bar{\mathbf{x}}_i^a] = E[\mathbf{x}_i^t - \mathbf{x}_i^a]$ . Then, from the linearization of  $\mathbf{h}$  at  $\mathbf{x}^g$ , (2.41) implies that

$$\mathbf{x}^t - \mathbf{x}^a \approx (\mathbf{x}^t - \mathbf{x}^g) - \mathbf{K}[\boldsymbol{\varepsilon}^o + \mathbf{H}^g(\mathbf{x}^t - \mathbf{x}^g)] \quad (3.17)$$

$$= [\mathbf{I} - \mathbf{K}\mathbf{H}^g][\mathbf{x}^t - \mathbf{x}^g] - \mathbf{K}\boldsymbol{\varepsilon}^o. \quad (3.18)$$

To continue, it is necessary to find the expected value of  $\mathbf{x}^t - \mathbf{x}^g$ . Looking at the differences  $\mathbf{x}_i^t - \mathbf{x}_i^g$  for each time, notice that

$$E[\mathbf{x}_0^t - \mathbf{x}_0^g] = E[\mathbf{x}_0^t - \mathbf{x}_0^b] = \mathbf{0} \quad (3.19)$$

and, if  $\mathbf{M}_1^g$  denotes the linearization of  $\mathcal{M}_1$  at  $\mathbf{x}_0^g$ ,

$$\mathbf{x}_1^t - \mathbf{x}_1^g \approx (\boldsymbol{\varepsilon}_1^q - \mathbf{q}_1) + \mathbf{M}_1^g(\mathbf{x}_0^t - \mathbf{x}_0^g). \quad (3.20)$$

The expected value of  $\mathbf{x}_1^t - \mathbf{x}_1^g$  is

$$\mathbb{E}[\mathbf{x}_1^t - \mathbf{x}_1^g] \approx \mathbf{q}_1^t - \mathbf{q}_1 = \Delta \mathbf{q}_1. \quad (3.21)$$

In a similar way, it can be seen that by linearizing the model  $\mathcal{M}_2$  at  $\mathbf{x}_1^g$ ,

$$\mathbf{x}_2^t - \mathbf{x}_2^g \approx (\boldsymbol{\varepsilon}_2^g - \mathbf{q}_2) + \mathbf{M}_2^g(\mathbf{x}_1^t - \mathbf{x}_1^g) \quad (3.22)$$

which has the expected value

$$\mathbb{E}[\mathbf{x}_2^t - \mathbf{x}_2^g] \approx \Delta \mathbf{q}_2 + \mathbf{M}_2^g \Delta \mathbf{q}_1. \quad (3.23)$$

Continuing for each time  $t_i$  in the data assimilation window, it can be seen that

$$\mathbf{x}_N^t - \mathbf{x}_N^g \approx \sum_{j=1}^N \mathbf{M}_{j \rightarrow N}^g (\boldsymbol{\varepsilon}_j^g - \mathbf{q}_j) + \mathbf{M}_{0 \rightarrow N}^g (\mathbf{x}_0^t - \mathbf{x}_0^b) \quad (3.24)$$

has the expected value

$$\mathbb{E}[\mathbf{x}_N^t - \mathbf{x}_N^g] \approx \sum_{j=1}^N \mathbf{M}_{j \rightarrow N}^g \Delta \mathbf{q}_j. \quad (3.25)$$

By writing this in block matrix form, it is possible to see that

$$\mathbb{E}[\mathbf{x}^t - \mathbf{x}^g] \approx \begin{bmatrix} \mathbf{I} & & & & \\ \mathbf{M}_1^g & \mathbf{I} & & & \\ \mathbf{M}_{0 \rightarrow 2}^g & \mathbf{M}_2^g & \mathbf{I} & & \\ \vdots & \vdots & \ddots & \ddots & \\ \mathbf{M}_{0 \rightarrow N}^g & \mathbf{M}_{1 \rightarrow N}^g & \cdots & \mathbf{M}_N^g & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \Delta \mathbf{q}_1 \\ \Delta \mathbf{q}_2 \\ \vdots \\ \Delta \mathbf{q}_N \end{bmatrix}, \quad (3.26)$$

From (3.18) and (3.26), it is possible to determine that

$$\mathbb{E}[\mathbf{x}^t - \mathbf{x}^a] \approx [\mathbf{I} - \mathbf{K}\mathbf{H}^g]\mathbf{F}^g\Delta\mathbf{q}, \quad (3.27)$$

where

$$\Delta\mathbf{q} = \begin{bmatrix} \mathbf{0}^T & \Delta\mathbf{q}_1^T & \cdots & \Delta\mathbf{q}_N^T \end{bmatrix}^T, \quad (3.28)$$

implying that the analysis states  $\mathbf{x}_i^a$  are unbiased estimates of the true states  $\mathbf{x}_i^t$  only if  $\Delta\mathbf{q} = \mathbf{0}$ , that is, when the background and model errors are unbiased. Then, by (3.14),

$$\boldsymbol{\varepsilon}^q - \mathbf{q}_e \approx \begin{bmatrix} -\mathbf{M}_1^a & \mathbf{I} & & & \\ & -\mathbf{M}_2^a & \mathbf{I} & & \\ & & \ddots & \ddots & \\ & & & -\mathbf{M}_N^a & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}_0^t - \bar{\mathbf{x}}_0^a \\ \mathbf{x}_1^t - \bar{\mathbf{x}}_1^a \\ \vdots \\ \mathbf{x}_N^t - \bar{\mathbf{x}}_N^a \end{bmatrix}. \quad (3.29)$$

If we denote the matrix in (3.29) by  $\tilde{\mathbf{F}}^a$ , then

$$\mathbb{E}[\boldsymbol{\varepsilon}^q - \mathbf{q}_e] \approx \tilde{\mathbf{F}}^a[\mathbf{I} - \mathbf{K}\mathbf{H}^g]\mathbf{F}^g\Delta\mathbf{q}. \quad (3.30)$$

If all of the error biases  $\mathbf{q}_i$  are specified as the true error biases  $\mathbf{q}_i^t$ , in which case  $\Delta\mathbf{q} = \mathbf{0}$ , then the ensemble model error average is unbiased.

### 3.4 Numerical Results

Preliminary numerical results are presented using the Lorenz models as described in Section 1.4. A data assimilation window consists of the current time  $t_0$  and three

time-steps, representing an assimilation window  $[t_0, t_3]$ . Observational data are generated from the true state with the observational error taken from the distribution  $N(0, (\sigma^o)^2)$  with the standard deviation specified as  $\sigma^o = 0.55$ . The observation operator satisfies  $\mathbf{h}_i(\mathbf{x}_i) = \mathbf{x}_i$  for  $i = 1, 2, 3$ .

An analysis will be produced from w4D-Var after setting up the background error covariance  $\mathbf{B}$  by running the extended Kalman filter using the true model error statistics for 700 time-steps with  $\mathbf{B}$  initialized to the identity matrix. The background  $\mathbf{x}_0^b$  for each step of the extended Kalman filter is taken to be a forecast of the previous analysis perturbed by random noise. After the spin-up cycle is complete,  $\mathbf{B}$  will then remain static for w4D-Var assimilation.

A comparative analysis is done to investigate the performance of the ensemble and hybrid assimilation methods to gauge their benefits. Three w4D-Var schemes (henceforth referred to as Control, Weak Ensemble, and Weak Hybrid) are run concurrently in order to properly compare and contrast the results. For each assimilation system, the background and observation error covariances are specified as described above, whereas the model error statistics are set as follows.

- (Control) Mis-specified model error covariances specified as  $\mathbf{Q} = 2 \text{diag}(\mathbf{Q}^t)$  and model bias  $\mathbf{q} = \mathbf{0}$  is considered as the status-quo and serves as the basis for comparing against the other schemes.
- (Weak Ensemble) Use equations (3.6) and (3.8) from an ensemble size of 20. The background was perturbed using (3.2) and multiplicative factor  $\beta = 10$ , which was used to make sure the ensemble had sufficient spread. The localization matrix is obtained using the fifth-order rational function of Gaspari and Cohn (3.9) with  $c = 8$ .
- (Weak Hybrid) Use equations (3.11) and (3.12) with  $\alpha = 0.5$ . The static com-

ponents are set to  $\mathbf{Q}_c = 2 \text{diag}(\mathbf{Q}^t)$  and  $\mathbf{q}_c = \mathbf{0}$  and the ensemble covariance with localization is the same one computed for the weak ensemble scheme.

Additionally, three strong-constraint 4D-Var schemes are considered. With the same observation error covariances as the w4D-Var systems, the background error covariance matrix is set as follows.

- (Strong 4D-Var) Set  $\mathbf{B}$  as the matrix from the spin-up cycle.
- (Strong Ensemble) Set  $\mathbf{B}$  as an ensemble estimate obtained from an ensemble of 20 strong-constraint 4D-Var assimilation systems. Further details are in the next paragraph.
- (Strong Hybrid) Use the hybrid  $\mathbf{B} = \alpha \mathbf{B}_c + (1 - \alpha) \mathbf{B}_e$ , where the static component  $\mathbf{B}_c$  is the covariance produced from the spin-up cycle and  $\mathbf{B}_e$  is the background ensemble error covariance with localization computed for the strong ensemble scheme. The parameter  $\alpha$  is also set to 0.5 like the weak hybrid scheme.

The ensemble-based schemes for the strong-constraint 4D-Var utilize background perturbations computed using (3.2) with multiplicative factor  $\beta = 5$ . In this case, the analysis  $\mathbf{x}_0^a$  in (3.2) is obtained from a strong-constraint 4D-Var run. The observations  $\mathbf{y}_i$  are perturbed in the same manner as described in Section 3.1. Each analysis ensemble member  $\mathbf{x}_{0,j}^a$  is forecast to the beginning of the next assimilation cycle and  $\mathbf{B}_e$  is defined to be the sample covariance from the ensemble of these forecasts.

To better compare the performance of the control system to the other five assimilation schemes, the ratio of errors between the five other schemes and the control is considered. For example, if the weak ensemble scheme outperforms the control scheme, then the ratio of the weak ensemble errors to the control errors will be less

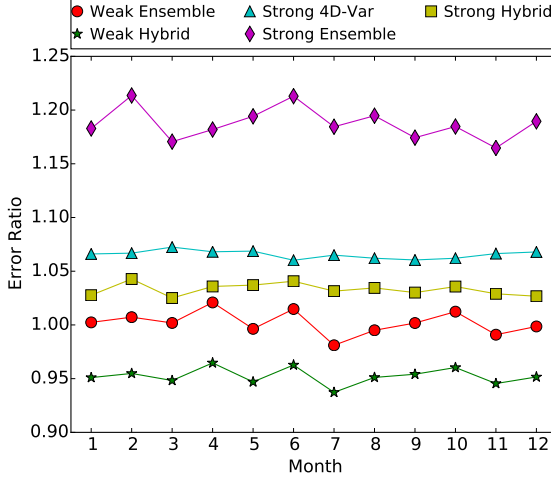


Figure 3.1: The ratio of the global monthly analysis errors to the control.

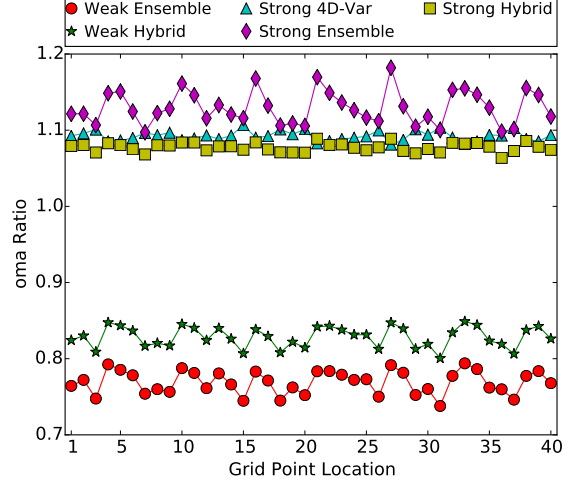


Figure 3.2: The ratio of the average oma differences to the control.

than 1. A ratio of 1 means the schemes have the same performance and larger than 1 if the weak ensemble scheme performs worse than the control.

Figure 3.1 shows the ratio of the monthly (30-day) average analysis errors for each method to the control, that is,  $\varepsilon^a(\langle \text{method} \rangle) / \varepsilon^a(\text{control})$ , where the analysis error is the difference between the LZ96  $\hat{x}$ -values and the w4D-Var analysis. It is noticed that the strong hybrid and weak ensemble schemes have similar performance, whose ratios fluctuate near 1. The weak ensemble scheme does not seem to perform much better than the control scheme, however the weak hybrid errors show an improvement over the entire assimilation period. At month 7, an improvement of about 7% is achieved. The fact that the ensemble scheme ratios are sometimes slightly larger than 1 can be attributed to two important components: the factor  $\beta$  that controls the background ensemble spread and the ensemble size. The choice to set  $\beta = 10$  for w4D-Var ensemble schemes was made to ensure that  $\mathbf{Q}_e$  did not suffer from being orders of magnitude smaller than  $\mathbf{Q}^t$ . Due to the banded structure of the localization matrix, the model error correlations are not fully accounted for in the ensemble and hybrid methods. Still, enough of the correlation structure was recovered from the ensembles

to reduce the monthly hybrid w4D-Var error averages.

Figure 3.2 shows the ratio of the three-year averaged observed-minus-analysis (oma)  $\|\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i^a)\|$  difference for each grid point to the control. It shows that the analyses for the weak ensemble scheme better fit to the observations than the hybrid scheme, even though Figure 3.1 shows the hybrid scheme had a lower average analysis error. Since the background, observational, and model error components of w4D-Var are weighted by their corresponding inverse covariance matrices in the cost functional, the analysis fit to the observations is affected by the magnitudes of the error covariance matrices. In particular,  $\mathbf{Q}_e$  having a larger magnitude than the hybrid model error covariance reduces the weight of model error in the analysis and increases the relative weights of the background and observations. Recalling that  $\mathbf{B}$  and  $\mathbf{R}$  remain unchanged between the two schemes, it can be inferred that  $\mathbf{Q}_e$  has a larger magnitude and that the hybrid specification better represents the true model error statistics. Some evidence to support this conjecture is shown in Figure 3.3, which compares the prescribed model error variance to the ensemble and hybrid model error variances, obtained from the three-year average covariance matrices.

The performance of a hybrid data assimilation system is closely determined by the weight assigned to the static and ensemble-based components of the error covariances. This aspect is investigated by running the hybrid data assimilation scheme for different  $\alpha$  for  $\mathbf{Q} = \alpha\mathbf{Q}_c + (1 - \alpha)\mathbf{Q}_e \circ \mathbf{C}$ , where  $\mathbf{Q}_c$  is the static component of  $\mathbf{Q}$  specified as the control error covariance. For  $\alpha = 0$ , the system runs in ensemble mode while for  $\alpha = 1$ , the system runs as the control, the status quo. The weight  $\alpha$  varies from 0 to 1 in steps of  $\Delta\alpha = 0.025$  and the ratio of the time- and space-averaged analysis errors over a three-year period to the control verses the choice of  $\alpha$  is shown in Figure 3.4. The ensemble size for generating  $\mathbf{Q}_e$  is 20, as before. The results show that the error corresponding to pure ensemble mode provides an improvement over the control

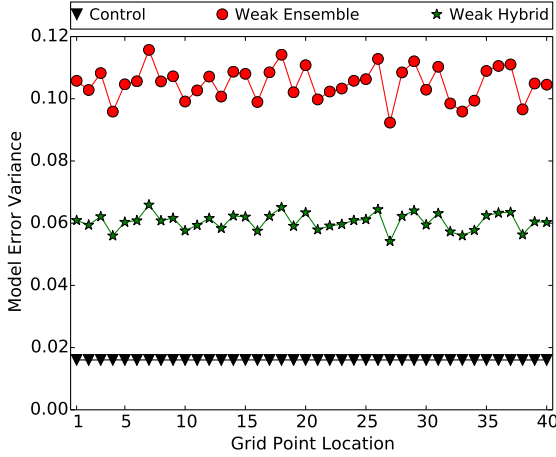


Figure 3.3: Comparison of the prescribed model error variance to the time-averaged ensemble and hybrid model error variances.

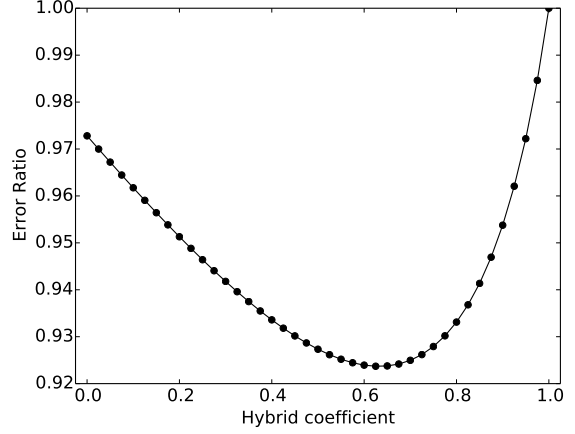


Figure 3.4: The ratio of the time- and space-averaged analysis errors to the control versus the hybrid scalar weight  $\alpha$  for the model error covariances.

data assimilation system by about 2.5%. Further reduction in the analysis error is achieved due to the specification of the hybrid covariance with  $0 < \alpha \leq 0.95$ . In particular, the hybrid covariance matrix corresponding to approximately  $\alpha = 0.625$  provides the greatest reduction in the analysis error, about 7.5% improvement over the control.

### 3.5 Conclusion

This chapter provides a framework for performing ensemble and hybrid data assimilation in a w4D-Var setting. A practical approach is considered that relies on an ensemble of w4D-Var systems solved by the incremental algorithm to obtain an ensemble of analysis sequences, the best estimates of the true state from which an ensemble of model error estimates is formed. These model error ensembles provide insight to the true nature of the model error covariance matrices. Model error bias has traditionally been assumed to be zero, and it may be possible to incorporate information about model error bias to improve the quality of the analysis with future



research in this area.

In some situations, such as the case when the number of ensemble members is small, the ensemble covariance matrices will have low rank and may not be a completely reliable representation of the true model error statistics. The weighted combination of a static matrix, a diagonal matrix, for example, and the ensemble covariance can prove to be an improvement over the ensemble matrices alone. A further improvement is to remove the random noise within the ensemble covariance by using a localization matrix.

The results of these numerical experiments provide a proof-of-concept for using ensembles in a w4D-Var setting. Specifying the model error covariances as the ensemble covariances with localization can improve the analysis error. Further improvement can be made in a hybrid setting with a good choice of the scalar weights.

The numerical results in this study assumed an idealized observing system in which all states in the dynamical system are observed. In practical applications, the performance of the data assimilation system is closely determined by the observing system configuration and further research is needed to investigate the performance of both ensemble and hybrid w4D-Var assimilation schemes.

## Chapter 4

### Observation Space Diagnostics

While the true error statistics may be unknown, a potential source of information may lie in observation residuals, the difference between observations and the observation space equivalent of the state vector, such as the observed-minus-background and observed-minus-analysis differences. Observation residuals have previously been used to estimate error statistics in data assimilation, such as the work done by Hollingsworth and Lönnberg [45], in which the observed-minus-background difference is used to estimate the background error covariance, and Desroziers and Ivanov [28], who used an a posteriori diagnosis with observed-minus-analysis differences. Dee has used observed-minus-background residuals to detect bias in data assimilation [21, 22]. Additionally, observed-minus-background residuals have been used to tune parametrizations of the background and observation error covariances (Chapnik et al. [9] for example).

In this chapter, the use of observation residuals as a diagnostic tool in w4D-Var is proposed, extending the diagnostics of Desroziers et al. [27] from one time-step to a data assimilation window  $[t_0, t_N]$ . The diagnostics of Desroziers et al. have previously been expanded upon by Todling [77] to estimate the model error covariance matrix using a lag-1 smoother and further to the variational case [76], but not a w4D-Var system.

## 4.1 Derivation of Diagnostics

To help simplify the following derivations, it will be assumed that the true model error biases are all zero. This means that the guess states satisfy

$$\mathbf{x}_0^g = \mathbf{x}_0^b \quad (4.1a)$$

$$\mathbf{x}_i^g = \mathcal{M}_i(\mathbf{x}_{i-1}^g), \quad i = 1, \dots, N \quad (4.1b)$$

implying that both the guess vector  $\mathbf{g}$  and bias vector  $\mathbf{q}$  from the incremental method are zero. With these assumptions, the w4D-Var analysis satisfies (2.41). If the error statistics are mis-specified in the DAS, then the analysis is still given by (2.41), but the gain matrix will be suboptimal, i.e.  $\mathbf{K} \neq \mathbf{K}^t$ , where  $\mathbf{K}^t$  is the gain matrix (2.38) corresponding to true error covariances.

Before the analysis  $\mathbf{x}^a$  is found, an a priori diagnostic can be formed using the observed-minus-guess difference  $\mathbf{d}_g^o = \mathbf{y} - \mathbf{h}(\mathbf{x}^g)$ , the same as (2.24) but renamed for consistency of notation in this section. From the linearization of the four-dimensional observation operator  $\mathbf{h}$  at  $\mathbf{x}^g$ ,

$$\mathbf{d}_g^o = \mathbf{y} - \mathbf{h}(\mathbf{x}^g) \approx \boldsymbol{\varepsilon}^o + \mathbf{H}^g(\mathbf{x}^t - \mathbf{x}^g). \quad (4.2)$$

If it is assumed that the background error  $\boldsymbol{\varepsilon}^b$  is uncorrelated with the model errors  $\boldsymbol{\varepsilon}_i^q$  and, in addition, the  $\boldsymbol{\varepsilon}_i^q$  are uncorrelated in time, then  $\mathbf{P}^t$  from (2.27) is the covariance matrix of

$$\boldsymbol{\varepsilon}^p = \begin{bmatrix} (\boldsymbol{\varepsilon}^b)^T & (\boldsymbol{\varepsilon}_1^q)^T & \dots & (\boldsymbol{\varepsilon}_N^q)^T \end{bmatrix}^T. \quad (4.3)$$

From linearizations of the model  $\mathcal{M}_i$  centered at  $\mathbf{x}_{i-1}^g$  as defined by (4.1), it can be

seen that  $\mathbf{x}^t - \mathbf{x}^g \approx \mathbf{F}^g \boldsymbol{\varepsilon}^p$ , where  $\mathbf{F}^g$  is from (2.30), by writing out the expression  $\mathbf{x}_i^t - \mathbf{x}_i^g$  for each  $i = 0, 1, \dots, N$ . First,

$$\mathbf{x}_0^t - \mathbf{x}_0^g = \boldsymbol{\varepsilon}^b \quad (4.4)$$

since  $\mathbf{x}_0^g = \mathbf{x}_0^b$ . Next,

$$\begin{aligned} \mathbf{x}_1^t - \mathbf{x}_1^g &= \mathbf{x}_1^t - \mathcal{M}_1(\mathbf{x}_0^t) + \mathcal{M}_1(\mathbf{x}_0^t) - \mathcal{M}_1(\mathbf{x}_0^b) \\ &\approx \boldsymbol{\varepsilon}_1^q + \mathbf{M}_1^g \boldsymbol{\varepsilon}^b \end{aligned} \quad (4.5)$$

where  $\mathbf{M}_1^g$  is evaluated at  $\mathbf{x}_0^g = \mathbf{x}_0^b$ . Then

$$\begin{aligned} \mathbf{x}_2^t - \mathbf{x}_2^g &= \mathbf{x}_2^t - \mathcal{M}_2(\mathbf{x}_1^t) + \mathcal{M}_2(\mathbf{x}_1^t) - \mathcal{M}_{0 \rightarrow 2}(\mathbf{x}_0^b) \\ &\approx \boldsymbol{\varepsilon}_2^q + \mathbf{M}_2^g(\mathbf{x}_1^t - \mathcal{M}_1(\mathbf{x}_0^b)) \\ &\approx \boldsymbol{\varepsilon}_2^q + \mathbf{M}_2^g \boldsymbol{\varepsilon}_1^q + \mathbf{M}_{0 \rightarrow 2}^g \boldsymbol{\varepsilon}^b \end{aligned} \quad (4.6)$$

where  $\mathbf{M}_2^g$  is evaluated at  $\mathbf{x}_1^g$ . Continuing in this way, it is noticed that

$$\mathbf{x}_N^t - \mathbf{x}_N^g \approx \sum_{j=1}^N \mathbf{M}_{j \rightarrow N}^g \boldsymbol{\varepsilon}_j^q + \mathbf{M}_{0 \rightarrow N}^g \boldsymbol{\varepsilon}^b, \quad (4.7)$$

where  $\mathbf{M}_{j \rightarrow N}^g$  is defined as in (2.31). By writing all of these guess errors  $\mathbf{x}_i^t - \mathbf{x}_i^g$  in block matrix form, it can be seen that  $\mathbf{x}^t - \mathbf{x}^g \approx \mathbf{F}^g \boldsymbol{\varepsilon}^p$  indeed holds. Hence,

$$\mathbf{d}_g^o \approx \boldsymbol{\varepsilon}^o + \mathbf{H}^g \mathbf{F}^g \boldsymbol{\varepsilon}^p. \quad (4.8)$$

If, it is further assumed that the observation errors  $\boldsymbol{\varepsilon}^o$  are uncorrelated with the four-dimensional error  $\boldsymbol{\varepsilon}^p$ , i.e.  $\boldsymbol{\varepsilon}_0^o$  is uncorrelated with  $\boldsymbol{\varepsilon}^p$  and  $\boldsymbol{\varepsilon}_i^o$  is uncorrelated with  $\boldsymbol{\varepsilon}_i^q$  for

$i = 1, \dots, N$ , then

$$\mathbf{E} [(\mathbf{d}_g^o)(\mathbf{d}_g^o)^T] \approx \mathbf{H}^g \mathbf{F}^g \mathbf{P}^t (\mathbf{F}^g)^T (\mathbf{H}^g)^T + \mathbf{R}^t. \quad (4.9)$$

Equation (4.9) can be used as a consistency check on the specification of  $\mathbf{R}$  and  $\mathbf{P}$  in that if  $\mathbf{R} \neq \mathbf{R}^t$  or

$$\mathbf{H}^g \mathbf{F}^g \mathbf{P} (\mathbf{F}^g)^T (\mathbf{H}^g)^T \neq (\mathbf{H}^g) \mathbf{F}^g \mathbf{P}^t (\mathbf{F}^g)^T (\mathbf{H}^g)^T, \quad (4.10)$$

that is, the observation error covariance is not correctly specified or the matrix  $\mathbf{F}^g \mathbf{P} (\mathbf{F}^g)^T$  does not agree with  $\mathbf{F}^g \mathbf{P}^t (\mathbf{F}^g)^T$  in observation space, then

$$\mathbf{E} [(\mathbf{d}_g^o)(\mathbf{d}_g^o)^T] = \mathbf{H}^g \mathbf{F}^g \mathbf{P} (\mathbf{F}^g)^T (\mathbf{H}^g)^T + \mathbf{R} \quad (4.11)$$

may not be true. Also, (4.9) resembles and reduces to equation (1) of Desroziers et al. [27] when the data assimilation window consists only of one time level  $t_0$ , as  $\mathbf{P}$  becomes  $\mathbf{B}$  and  $\mathbf{F}^g$  becomes the identity in this case. When the forecast model and observation operator are linear, (4.9) is an equality.

Now, the difference between  $\mathbf{x}^a$  and  $\mathbf{x}^g$  in observation space is

$$\mathbf{d}_g^a = \mathbf{h}(\mathbf{x}^a) - \mathbf{h}(\mathbf{x}^g) \approx \mathbf{H}^g (\mathbf{x}^a - \mathbf{x}^g) = \mathbf{H}^g \mathbf{K} \mathbf{d}_g^o, \quad (4.12)$$

which follows from (2.41). Thus

$$\mathbf{E} [(\mathbf{d}_g^a)(\mathbf{d}_g^a)^T] \approx \mathbf{H}^g \mathbf{K} \mathbf{E} [(\mathbf{d}_g^o)(\mathbf{d}_g^o)^T] \quad (4.13)$$

$$= \mathbf{H}^g \mathbf{K} [\mathbf{H}^g \mathbf{F}^g \mathbf{P}^t (\mathbf{F}^g)^T (\mathbf{H}^g)^T + \mathbf{R}^t]. \quad (4.14)$$

From the definition of  $\mathbf{K}$  in (2.38),

$$\begin{aligned} \mathbb{E} [(\mathbf{d}_g^a)(\mathbf{d}_g^o)^T] &\approx \mathbf{H}^g \mathbf{F}^g \mathbf{P} (\mathbf{F}^g)^T (\mathbf{H}^g)^T [\mathbf{H}^g \mathbf{F}^g \mathbf{P} (\mathbf{F}^g)^T (\mathbf{H}^g)^T + \mathbf{R}]^{-1} \\ &\times [\mathbf{H}^g \mathbf{F}^g \mathbf{P}^t (\mathbf{F}^g)^T (\mathbf{H}^g)^T + \mathbf{R}^t] \end{aligned} \quad (4.15)$$

and when  $\mathbf{H}^g \mathbf{K} = \mathbf{H}^g \mathbf{K}^t$ , (4.15) reduces to

$$\mathbb{E} [(\mathbf{d}_g^a)(\mathbf{d}_g^o)^T] \approx \mathbf{H}^g \mathbf{F}^g \mathbf{P}^t (\mathbf{F}^g)^T (\mathbf{H}^g)^T. \quad (4.16)$$

This result is in agreement with (2) with Desroziers et al. [27] when the data assimilation window consists of only one time level,  $t_0$ . Like before, if  $\mathbf{h}_i$  and  $\mathcal{M}_i$  are linear, (4.16) is an equality.

Similarly, the observed-minus-analysis difference  $\mathbf{d}_a^o$  is

$$\mathbf{d}_a^o = \mathbf{y} - \mathbf{h}(\mathbf{x}^a) \approx \mathbf{d}_g^o - \mathbf{H}^g \mathbf{K} \mathbf{d}_g^o = (\mathbf{I} - \mathbf{H}^g \mathbf{K}) \mathbf{d}_g^o \quad (4.17)$$

It can be verified that

$$\mathbf{I} - \mathbf{H}^g \mathbf{K} = \mathbf{R} [\mathbf{H}^g \mathbf{F}^g \mathbf{P} (\mathbf{F}^g)^T (\mathbf{H}^g)^T + \mathbf{R}]^{-1} \quad (4.18)$$

and so, from (4.9),

$$\begin{aligned} \mathbb{E} [(\mathbf{d}_a^o)(\mathbf{d}_g^o)^T] &\approx \mathbf{R} [\mathbf{H}^g \mathbf{F}^g \mathbf{P} (\mathbf{F}^g)^T (\mathbf{H}^g)^T + \mathbf{R}]^{-1} \\ &\times [\mathbf{H}^g \mathbf{F}^g \mathbf{P}^t (\mathbf{F}^g)^T (\mathbf{H}^g)^T + \mathbf{R}^t]. \end{aligned} \quad (4.19)$$

Once again, if the specification of  $\mathbf{H}^g \mathbf{K}$  agrees with  $\mathbf{H}^g \mathbf{K}^t$ , then  $\mathbf{I} - \mathbf{H}^g \mathbf{K}$  from (4.18)

equals

$$\mathbf{I} - \mathbf{H}^g \mathbf{K}^t = \mathbf{R}^t [\mathbf{H}^g \mathbf{F}^g \mathbf{P}^t (\mathbf{F}^g)^T (\mathbf{H}^g)^T + \mathbf{R}^t]^{-1} \quad (4.20)$$

and (4.19) can be further simplified to

$$\mathbf{E} [(\mathbf{d}_a^o)(\mathbf{d}_g^o)^T] \approx \mathbf{R}^t, \quad (4.21)$$

analogous to equation (3) of Desroziers et al. [27]. Due to the block diagonal structure of  $\mathbf{R}^t$ , this readily translates into a statement about the individual components of  $\mathbf{d}_a^o$  and  $\mathbf{d}_g^o$ . Firstly,

$$\mathbf{E} [(\mathbf{d}_a^o)_i (\mathbf{d}_g^o)_i^T] \approx \mathbf{R}_i^t, \quad (4.22)$$

and secondly,

$$\mathbf{E} [(\mathbf{d}_a^o)_i (\mathbf{d}_g^o)_j^T] \approx \mathbf{0} \quad (4.23)$$

when  $i \neq j$ . Therefore, the components of  $\mathbf{d}_a^o$  and  $\mathbf{d}_g^o$  are uncorrelated in time.

The final diagnostic formulated here is the diagnosis for the analysis errors, similar to equation (4) of Desroziers et al. [27]. Consider the product

$$\mathbf{d}_g^a (\mathbf{d}_a^o)^T \approx \mathbf{H}^g \mathbf{K} (\mathbf{d}_g^o) (\mathbf{d}_g^o)^T [\mathbf{H}^g \mathbf{F}^g \mathbf{P}^t (\mathbf{F}^g)^T (\mathbf{H}^g)^T + \mathbf{R}^t]^{-1} \mathbf{R}. \quad (4.24)$$

By (4.9) again,

$$\begin{aligned} \mathbf{E} [(\mathbf{d}_g^a)(\mathbf{d}_a^o)^T] &\approx \mathbf{H}^g \mathbf{K} [\mathbf{H}^g \mathbf{F}^g \mathbf{P}^t (\mathbf{F}^g)^T \mathbf{H}^T + \mathbf{R}^t] \\ &\times [\mathbf{H}^g \mathbf{F}^g \mathbf{P} (\mathbf{F}^g)^T (\mathbf{H}^g)^T + \mathbf{R}]^{-1} \mathbf{R}. \end{aligned} \quad (4.25)$$

When  $\mathbf{H}^g \mathbf{K} = \mathbf{H}^g \mathbf{K}^t$ , this expression reduces to

$$\mathbf{E} [(\mathbf{d}_g^a)(\mathbf{d}_a^o)^T] \approx \mathbf{H}^g \mathbf{K}^t \mathbf{R}^t. \quad (4.26)$$

All of these diagnostics provide consistency checks for a linear data assimilation system. In the case of w4D-Var, if the model and observation operators are weakly nonlinear, these diagnostics are an approximation and should still provide meaningful information. For the case when assimilation is only performed at one time level, all the results presented here reduce to their three-dimensional counterparts derived by Desroziers et al. [27].

## 4.2 A Special Case

The diagnostic equations presented in the previous section do not provide information on the full matrix structure of  $\mathbf{P}^t$  unless observations are available at all grid points, that is,  $\mathbf{h}_i(\mathbf{x}_i) = \mathbf{x}_i$ . In this case, the four-dimensional operator  $\mathbf{H}^g$  is the identity matrix and the diagnostics (4.9) and (4.16) can then be expressed as

$$\mathbf{E} [(\mathbf{d}_g^o)(\mathbf{d}_g^o)^T] = \mathbf{F}^g \mathbf{P}^t (\mathbf{F}^g)^T + \mathbf{R}^t \quad (4.27)$$

$$\mathbf{E} [(\mathbf{d}_g^a)(\mathbf{d}_g^o)^T] = \mathbf{F}^g \mathbf{P}^t (\mathbf{F}^g)^T. \quad (4.28)$$



By isolating  $\mathbf{P}^t$ ,

$$\mathbf{P}^t = \mathbf{E} [[(\mathbf{F}^g)^{-1} \mathbf{d}_g^o][(\mathbf{F}^g)^{-1} \mathbf{d}_g^o]^T] - (\mathbf{F}^g)^{-1} \mathbf{R}^t (\mathbf{F}^g)^{-T} \quad (4.29)$$

$$\mathbf{P}^t = \mathbf{E} [[(\mathbf{F}^g)^{-1} \mathbf{d}_g^a][(\mathbf{F}^g)^{-1} \mathbf{d}_g^o]^T], \quad (4.30)$$

equations (4.29) and (4.30) serve as a priori and a posteriori diagnostics, respectively. When  $\mathbf{P}^t$  is constant in time, the expectations in (4.29) and (4.30) can be estimated by the time average of rank one matrices of the form  $[(\mathbf{F}^g)^{-1} \mathbf{d}_g^o][(\mathbf{F}^g)^{-1} \mathbf{d}_g^o]^T$  and  $[(\mathbf{F}^g)^{-1} \mathbf{d}_g^a][(\mathbf{F}^g)^{-1} \mathbf{d}_g^o]^T$  over a number of data assimilation cycles. The special structure of  $(\mathbf{F}^g)^{-1}$  can be exploited so that the product of  $(\mathbf{F}^g)^{-1}$  with a block vector  $\mathbf{d}$

$$\begin{bmatrix} \mathbf{I} & & & & \\ -\mathbf{M}_1^g & \mathbf{I} & & & \\ & & -\mathbf{M}_2^g & \mathbf{I} & \\ & & & \ddots & \ddots \\ & & & & -\mathbf{M}_N^g & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{d}_0 \\ \mathbf{d}_1 \\ \mathbf{d}_2 \\ \vdots \\ \mathbf{d}_N \end{bmatrix} = \begin{bmatrix} \mathbf{d}_0 \\ \mathbf{d}_1 - \mathbf{M}_1^g \mathbf{d}_0 \\ \mathbf{d}_2 - \mathbf{M}_2^g \mathbf{d}_1 \\ \vdots \\ \mathbf{d}_N - \mathbf{M}_N^g \mathbf{d}_{N-1} \end{bmatrix} \quad (4.31)$$

only requires the use of the tangent linear models and vector subtraction. Since the guess states for the incremental method will be different from cycle to cycle, the matrix  $\mathbf{F}^g$  will also be different. Similarly, if  $\mathbf{R}^t$  is constant in time, the expectation in (4.21) can be estimated by a time average of the rank one matrices  $(\mathbf{d}_a^o)(\mathbf{d}_g^o)^T$ .

The covariance matrices  $\mathbf{P}^t$  and  $\mathbf{R}^t$  are symmetric, however the time averages estimating the expected values in (4.21) and (4.30) are not symmetric. So, for practical purposes, their diagnostic estimates may be identified with their symmetric parts. For

any square matrix  $\mathbf{X}$ , the symmetric part of  $\mathbf{X}$  is the symmetric matrix  $\frac{1}{2}(\mathbf{X} + \mathbf{X}^T)$ .

$$\mathbf{P}^t \approx \frac{1}{2(\# \text{ cycles})} \sum [[(\mathbf{F}^g)^{-1} \mathbf{d}_g^a][(\mathbf{F}^g)^{-1} \mathbf{d}_g^o]^T + [(\mathbf{F}^g)^{-1} \mathbf{d}_g^o][(\mathbf{F}^g)^{-1} \mathbf{d}_g^a]^T] \quad (4.32)$$

$$\mathbf{R}^t \approx \frac{1}{2(\# \text{ cycles})} \sum [(\mathbf{d}_g^a)(\mathbf{d}_g^o)^T + (\mathbf{d}_g^o)(\mathbf{d}_g^a)^T] \quad (4.33)$$

The estimation of  $\mathbf{P}^t$  via (4.29) is symmetric, however, so

$$\mathbf{P}^t \approx \frac{1}{\# \text{ cycles}} \sum [[(\mathbf{F}^g)^{-1} \mathbf{d}_g^o][(\mathbf{F}^g)^{-1} \mathbf{d}_g^o]^T - (\mathbf{F}^g)^{-1} \mathbf{R}^t (\mathbf{F}^g)^{-T}] \quad (4.34)$$

will be sufficient in producing a symmetric approximation of  $\mathbf{P}^t$ . Any discrepancies that appear from comparing both sides of these approximations will indicate that the data assimilation system is suboptimal.

The first  $n \times n$  block of (4.32) and (4.34) are diagnostics for the background error covariance matrix  $\mathbf{B}^t$  and with the assumption that  $\mathbf{Q}_i^t$  is constant in time, the other  $n \times n$  diagonal blocks all estimate the time-invariant  $\mathbf{Q}_i^t$ , which can be averaged together for another estimate of the model error covariance. Similarly, the diagonal blocks of (4.33) can be averaged together for another estimate of the time-invariant  $\mathbf{R}_i^t$ .

### 4.3 Numerical Results

Once again, preliminary numerical results are presented for the Lorenz models from Section 1.4. This time, observational data generated from the true state, corrupted by random noise from  $N(0, (\sigma^o)^2)$  with the standard deviation specified as  $\sigma^o = 0.55$ , are available at every time in the assimilation window  $[t_0, t_3]$ . The observation operator is specified as  $\mathbf{h}_i(\mathbf{x}_i) = \mathbf{x}_i$  so that equation (4.32) can be used in order to get an estimate of  $\mathbf{Q}^t$ . The background error covariance is defined to be a

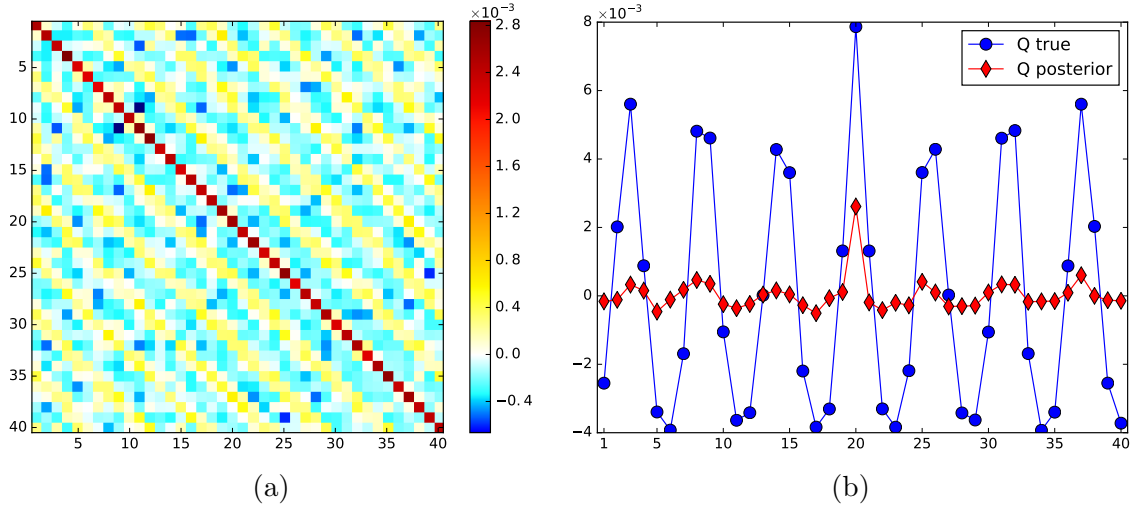


Figure 4.1: (a) Posterior estimate of  $\mathbf{Q}^t$ , (b) Twentieth rows of  $\mathbf{Q}^t$  and the posterior estimate.

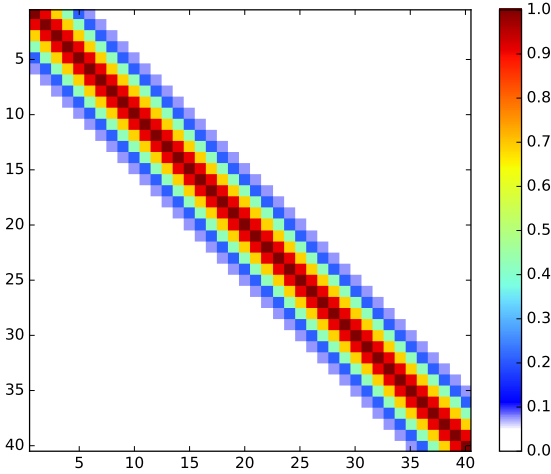


Figure 4.2: Specification of the background error correlation matrix.

static band matrix with correlation structure specified by the fifth-order rational function of compact support (3.9) with correlation parameter  $c = 4$ . The background error variance is set to  $(\sigma^b)^2 = 0.1$  for each grid point. Initially, the background  $\mathbf{x}_0^b$  is the truth perturbed by random noise and is a forecast of a previous analysis for future cycles. Figure 4.2 shows the specification of the background error correlation matrix.

The weak-constraint diagnostic matrices (4.32) and (4.33) are estimated over a one-year time period and compared against the true error covariance matrices. Since the model error covariance matrix is constant in time, the three  $n \times n$  diagonal blocks of the estimate from (4.32) that all estimate the same time-invariant  $\mathbf{Q}^t$  are averaged

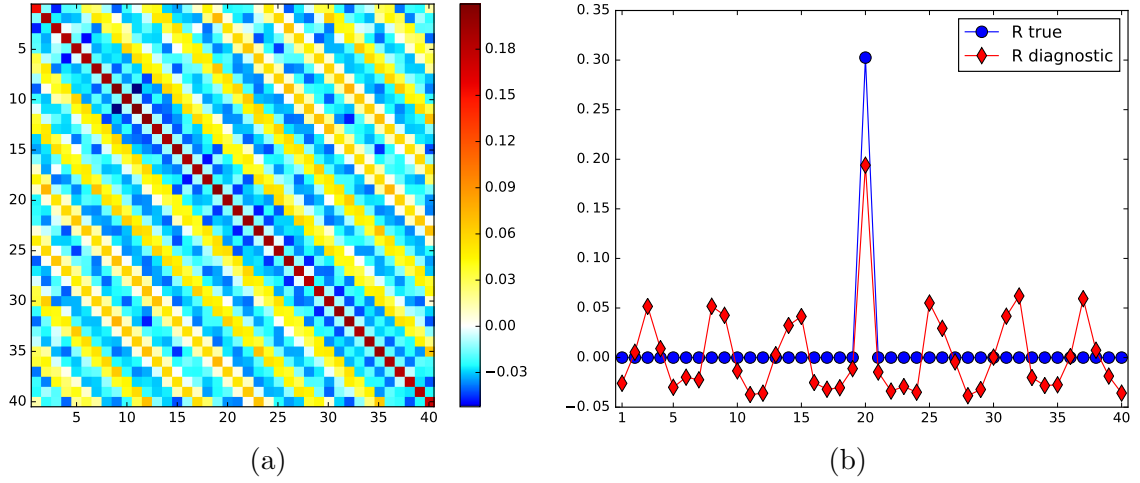


Figure 4.3: (a) Posterior estimate of  $\mathbf{R}^t$ , (b) Twentieth rows of  $\mathbf{R}^t$  and the posterior estimate.

together. Figure 4.1 shows the posterior  $\mathbf{Q}$ -diagnostic matrix from this experiment along with the comparison between the twentieth rows of  $\mathbf{Q}^t$  and the diagnostic matrix. By comparing the  $\mathbf{Q}$ -diagnostic matrix with the true model error covariance matrix for the Lorenz model from Figure 1.1b, it is clear that the  $\mathbf{Q}$ -diagnostic matrix has recovered the correlation structure of  $\mathbf{Q}^t$ . The magnitude of the diagnostic matrix, however, appears to be lower than the true model error covariance. While  $\mathbf{Q}^t$  was not recovered entirely, the posterior diagnostic matrix hints at the existence of model error correlations that were not accounted for in the specification of  $\mathbf{Q}$  in the DAS.

In a similar manner, the four diagonal blocks of the estimate from (4.33) are averaged together and compared to  $\mathbf{R}^t$ . Figure 4.3 shows the  $\mathbf{R}$ -diagnostic matrix along with a cross-sectional analysis of the twentieth rows of  $\mathbf{R}^t$  and the  $\mathbf{R}$ -diagnostic matrices. Recalling that  $\mathbf{R} = \mathbf{R}^t$  is a diagonal matrix, the true model error correlation structure has become incorporated into the observation error covariance estimate. This is verified by comparing the the rows of  $\mathbf{Q}^t$  and the  $\mathbf{R}$ -diagnostic matrix. The twentieth row of the  $\mathbf{R}$ -diagnostic matrix follows the same correlation pattern exhib-

ited by  $\mathbf{Q}^t$ . This phenomenon was also seen by Todling [76], who points out that this has consequences for how estimates of observation error correlations derived from observation residuals are interpreted in operational data assimilation procedures. If, in an operational DAS,  $\mathbf{R}$  has been accurately specified, the appearance of correlations within the  $\mathbf{R}$ -diagnostic matrix indicates a problem with the specification of  $\mathbf{Q}$ , i.e. that the model error correlations have been mis-specified.

#### 4.4 Conclusion

The work in this chapter provides a study on the framework of observation space diagnostics in a w4D-Var setting. These new diagnostic equations are an extension of those derived by Desroziers et al. [27] and when only one time level is considered, the w4D-Var diagnostics reduce to the Desroziers diagnostics. A practical approach is considered that relies on the evaluation of statistical expectations involving observation residuals, which contain some information about the true error statistics.

In an operational DAS, the statistical expectations can be estimated practically over a number of assimilation cycles from a single realization of the residuals from each cycle. While model error bias was assumed to be zero in the derivation of the diagnostic equations, a nonzero model error bias can be considered and with slight modifications to the work of Section 4.1, the same diagnostic matrices can be made.

Numerical experiments considered in this chapter demonstrated the ability of the diagnostics to recover information about the true model error covariance in a suboptimal DAS. More specifically, only  $\mathbf{Q}$  was mis-specified and the diagnostics attempted to recover the true model error covariance. The posterior  $\mathbf{Q}$ -diagnostic made it clear that there exist model error correlations that were absent in the DAS. The observation residuals included information about the true model error correlation structure and incorporated it into the estimate of  $\mathbf{R}^t$  that, at first glance, indicated

the presence of observation error correlations that did not exist.

## Chapter 5

### Sensitivity Analysis

The weighting between the information provided by models and observational data in a 4D-Var data assimilation system is determined by the representation of the statistical properties of the errors in the background, model, and observations. Adjoint-based sensitivity analysis allows for the development of efficient methodologies to assess the contribution of the various input parameters to the reduction of forecast error. Le Dimet and Ngodock [55] provide the general framework for performing sensitivity analysis in variational data assimilation, such as 4D-Var. Baker and Daley [4] derived the equations of the forecast sensitivity to observations and the background state in terms of the adjoint of the DAS. Subsequently, these techniques have been extended to incorporate the forecast error sensitivity to observation and background error covariances [14, 15, 16, 19, 20] and provided a basis for tuning error covariance parameters [26, 28]. Cioaca et al. [11] present a practical approach to quantify the impact associated with distinct observing system components. Akella and Navon [1] discuss a method for estimating scalar parameters in the representation of systematic model errors through variational data assimilation. Yaremchuk and Martin [81] compare 4D-Var ensemble sensitivity with adjoint sensitivity analysis for a quasi-geostrophic model. Lupu et al. [65] use sensitivities and innovation statistics to tune the observation error covariance matrix in numerical weather prediction. Cioaca and Sandu [10] develop an optimization framework to improve 4D-Var data assimilation

system performance. Lorenc and Marriott [62] discuss an adjoint-based approach for calculating observation impacts in global 4D-Var at the Met Office.

The work in this chapter extends the adjoint-based approach to sensitivity analysis to the w4D-Var framework and provides the equations to evaluate the forecast sensitivity with respect to parameters used to represent the model error covariance in the DAS. Theoretical aspects and sensitivity guidance to forecast error reduction through model error tuning procedures are presented.

## 5.1 Sensitivity Analysis

The analysis  $\mathbf{x}^a$  obtained from a w4D-Var assimilation system is determined by the input parameters  $\mathbf{x}^a = \mathbf{x}^a(\mathbf{x}_0^b, \mathbf{y}_i, \mathbf{B}, \mathbf{R}_i, \mathbf{Q}_i, \mathbf{q}_i)$  and varying these inputs can impact the quality of the analysis. In 4D-Var, the impact of variations in the error covariance models  $\mathbf{R}$  and  $\mathbf{B}$  is evaluated for a short-range forecast error measure typically defined as

$$e(\mathbf{x}_k^a) = (\mathbf{x}_f^a - \mathbf{x}_f^v)^T \mathbf{E} (\mathbf{x}_f^a - \mathbf{x}_f^v) \quad (5.1)$$

where  $\mathbf{x}_k^a$  is the analysis at time  $t_k$  and  $\mathbf{x}_f^a = \mathcal{M}_{t_k \rightarrow t_f}(\mathbf{x}_k^a)$  is the model forecast of  $\mathbf{x}_k^a$  to verification time  $t_f$ ,  $\mathbf{x}_f^v$  is the verifying analysis at time  $t_f$  serving as a proxy to the true state  $\mathbf{x}_f^t$ , and  $\mathbf{E}$  is a diagonal matrix of weights. The methodology of deriving the 4D-Var sensitivity equations in the context of nonlinear least-squares optimization [17, 18]

$$J(\mathbf{x}) = \frac{1}{2} [\mathbf{\Gamma}(\mathbf{x}) - \mathbf{z}]^T \mathbf{W}^{-1} [\mathbf{\Gamma}(\mathbf{x}) - \mathbf{z}] \quad (5.2)$$



can be extended to the w4D-Var cost functional (1.19) by defining the information vector as

$$\mathbf{z} = \begin{bmatrix} \mathbf{x}_0^b \\ \vdots \mathbf{q}_i \\ \vdots \mathbf{y}_i \end{bmatrix} \in \mathbb{R}^{M+p}, \quad (5.3)$$

where  $M = (N + 1)n$ , the nonlinear operator  $\mathbf{\Gamma} : \mathbb{R}^M \rightarrow \mathbb{R}^{M+p}$  which maps the four-dimensional state  $\mathbf{x}$  into information space

$$\mathbf{\Gamma}(\mathbf{x}) = \begin{bmatrix} \mathbf{x}_0 \\ \vdots \mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1}) \\ \vdots \mathbf{h}_i(\mathbf{x}_i) \end{bmatrix}, \quad (5.4)$$

and the block diagonal covariance model

$$\mathbf{W} = \begin{bmatrix} \mathbf{P} & \mathbf{0}_{M \times p} \\ \mathbf{0}_{p \times M} & \mathbf{R} \end{bmatrix} \in \mathbb{R}^{(M+p) \times (M+p)}. \quad (5.5)$$

In the work that follows, the sensitivity of a scalar forecast error functional to the information  $\mathbf{z}$  and the covariance model  $\mathbf{W}$  is derived in the general context when the forecast error aspect  $e(\mathbf{x}_0^a, \dots, \mathbf{x}_N^a)$  is a function of the time-distributed model state.

## 5.2 Sensitivity Equations

The calculus of variations provides a tool for deriving the sensitivity of a scalar functional  $e(\mathbf{x})$  to a parameter  $\mathbf{X} \in \mathbb{R}^{m \times n}$  via the first-order variation

$$\delta e = \langle \nabla_{\mathbf{X}} e(\mathbf{x}), \delta \mathbf{X} \rangle_{\mathbb{R}^{m \times n}} \quad (5.6)$$

where the inner product is defined in terms of the matrix trace operator as

$$\langle \mathbf{X}, \mathbf{Y} \rangle_{\mathbb{R}^{m \times n}} = \text{Tr}(\mathbf{X}\mathbf{Y}^T). \quad (5.7)$$

The first-order optimality condition  $\nabla_{\mathbf{x}} J(\mathbf{x}^a) = \mathbf{0}$ , or

$$\mathbf{\Gamma}_{\mathbf{x}}^T(\mathbf{x}^a) \mathbf{W}^{-1} [\mathbf{\Gamma}(\mathbf{x}^a) - \mathbf{z}] = \mathbf{0}, \quad (5.8)$$

where

$$\mathbf{\Gamma}_{\mathbf{x}}(\mathbf{x}^a) = \frac{\partial \mathbf{\Gamma}}{\partial \mathbf{x}}(\mathbf{x}^a) = \begin{bmatrix} (\mathbf{F}^a)^{-1} \\ \mathbf{H}^a \end{bmatrix} \in \mathbb{R}^{(M+p) \times M} \quad (5.9)$$

denotes the Jacobian of  $\mathbf{\Gamma}$  evaluated at  $\mathbf{x}^a$ , is used to establish the relation between the first-order variation  $\delta \mathbf{x}^a$  and the parameter variations  $\delta \mathbf{z}$  and  $\delta \mathbf{W}$ . The matrix  $\mathbf{F}^a$  is like (2.30), but evaluated at the analysis state. It is noticed that

$$[\nabla_{\mathbf{xx}}^2 J(\mathbf{x}^a)] \delta \mathbf{x}^a - \mathbf{\Gamma}_{\mathbf{x}}^T(\mathbf{x}^a) \mathbf{W}^{-1} \delta \mathbf{z} + \mathbf{\Gamma}_{\mathbf{x}}^T(\mathbf{x}^a) \delta(\mathbf{W}^{-1}) [\mathbf{\Gamma}(\mathbf{x}^a) - \mathbf{z}] = \mathbf{0}, \quad (5.10)$$

where  $\nabla_{\mathbf{xx}}^2 J(\mathbf{x}^a)$  is the Hessian matrix of the w4D-Var cost functional evaluated at  $\mathbf{x}^a$ ,  $\mathbf{W}$ , and  $\mathbf{z}$ . Using the identity

$$\delta(\mathbf{W}^{-1}) = -\mathbf{W}^{-1} [\delta \mathbf{W}] \mathbf{W}^{-1} \quad (5.11)$$

(5.10) can be rewritten as

$$[\nabla_{\mathbf{xx}}^2 J(\mathbf{x}^a)] \delta \mathbf{x}^a - \mathbf{\Gamma}_{\mathbf{x}}^T(\mathbf{x}^a) \mathbf{W}^{-1} \delta \mathbf{z} - \mathbf{\Gamma}_{\mathbf{x}}^T(\mathbf{x}^a) \mathbf{W}^{-1} [\delta \mathbf{W}] \mathbf{W}^{-1} [\mathbf{\Gamma}(\mathbf{x}^a) - \mathbf{z}] = \mathbf{0}. \quad (5.12)$$

Taking the inner product of (5.12) with an arbitrary vector  $\chi \in \mathbb{R}^M$  gives

$$\begin{aligned} \langle \chi, [\nabla_{\mathbf{x}\mathbf{x}}^2 J(\mathbf{x}^a)] \delta \mathbf{x}^a \rangle_{\mathbb{R}^M} &= \langle \chi, \Gamma_{\mathbf{x}}^T(\mathbf{x}^a) \mathbf{W}^{-1} [\delta \mathbf{W}] \mathbf{W}^{-1} [\Gamma(\mathbf{x}^a) - \mathbf{z}] \rangle_{\mathbb{R}^M} + \\ &\quad \langle \chi, \Gamma_{\mathbf{x}}^T(\mathbf{x}^a) \mathbf{W}^{-1} \delta \mathbf{z} \rangle_{\mathbb{R}^M}. \end{aligned} \quad (5.13)$$

This can be simplified using the fact that  $\nabla_{\mathbf{x}\mathbf{x}}^2 J(\mathbf{x}^a)$  and  $\mathbf{W}$  are symmetric matrices and with the aid of some linear algebra operations.

$$\begin{aligned} \langle [\nabla_{\mathbf{x}\mathbf{x}}^2 J(\mathbf{x}^a)] \chi, \delta \mathbf{x}^a \rangle_{\mathbb{R}^M} &= \langle \mathbf{W}^{-1} \Gamma_{\mathbf{x}}(\mathbf{x}^a) \chi [\Gamma(\mathbf{x}^a) - \mathbf{z}]^T \mathbf{W}^{-1}, \delta \mathbf{W} \rangle_{\mathbb{R}^{(M+p) \times (M+p)}} + \\ &\quad \langle \mathbf{W}^{-1} \Gamma_{\mathbf{x}}(\mathbf{x}^a) \chi, \delta \mathbf{z} \rangle_{\mathbb{R}^{M+p}} \end{aligned} \quad (5.14)$$

Let  $\chi$  denote the solution to

$$[\nabla_{\mathbf{x}\mathbf{x}}^2 J(\mathbf{x}^a)] \chi = \nabla_{\mathbf{x}} e(\mathbf{x}^a), \quad (5.15)$$

where  $\nabla_{\mathbf{x}} e(\mathbf{x}^a)$  is the dimension  $M$  forecast error gradient. Second-order derivative information about the Hessian of the cost functional (5.2) may be obtained through the development of a second-order adjoint model [54].

From equations (5.14) and (5.15), the first-order variation in the forecast error functional is

$$\delta e = \langle \mathbf{W}^{-1} \Gamma_{\mathbf{x}}(\mathbf{x}^a) \chi, \delta \mathbf{z} \rangle_{\mathbb{R}^{M+p}} + \langle \mathbf{W}^{-1} \Gamma_{\mathbf{x}}(\mathbf{x}^a) \chi [\Gamma(\mathbf{x}^a) - \mathbf{z}]^T \mathbf{W}^{-1}, \delta \mathbf{W} \rangle_{\mathbb{R}^{M+p}}. \quad (5.16)$$

From (5.6), the forecast sensitivity to the information vector  $\mathbf{z}$  is

$$\nabla_{\mathbf{z}} e(\mathbf{x}^a) = \mathbf{W}^{-1} \Gamma_{\mathbf{x}}(\mathbf{x}^a) \chi \in \mathbb{R}^{M+p} \quad (5.17)$$

and the forecast sensitivity to the covariance model  $\mathbf{W}$  is

$$\nabla_{\mathbf{W}}e(\mathbf{x}^a) = \mathbf{W}^{-1}\mathbf{\Gamma}_{\mathbf{x}}(\mathbf{x}^a)\boldsymbol{\chi}[\mathbf{\Gamma}(\mathbf{x}^a) - \mathbf{z}]^T\mathbf{W}^{-1} \in \mathbb{R}^{(M+p) \times (M+p)}. \quad (5.18)$$

The  $\mathbf{W}$ -sensitivity equation (5.18) can be rewritten as

$$\nabla_{\mathbf{W}}e(\mathbf{x}^a) = \nabla_{\mathbf{z}}e(\mathbf{x}^a)[\mathbf{W}^{-1}(\mathbf{\Gamma}(\mathbf{x}^a) - \mathbf{z})]^T. \quad (5.19)$$

By denoting

$$\mathbf{v}_1 = \nabla_{\mathbf{z}}e(\mathbf{x}^a) \quad (5.20a)$$

$$\mathbf{v}_2 = \mathbf{W}^{-1}[\mathbf{\Gamma}(\mathbf{x}^a) - \mathbf{z}], \quad (5.20b)$$

the  $\mathbf{W}$ -sensitivity matrix can be represented as

$$\nabla_{\mathbf{W}}e(\mathbf{x}^a) = \mathbf{v}_1\mathbf{v}_2^T, \quad (5.21)$$

where it is noticed that the  $\mathbf{W}$ -sensitivity matrix has rank one. Thus, evaluation and storage of only two vectors is required to contain the  $\mathbf{W}$ -sensitivity information.

### 5.3 Sensitivities with Matrix Decomposition

The error covariance model  $\mathbf{W}$  is a symmetric and positive definite matrix. For practical purposes, it is therefore necessary to constrain  $\delta\mathbf{W}$  so that  $\mathbf{W} + \delta\mathbf{W}$  is also symmetric and positive definite. A perturbation using the  $\mathbf{W}$ -sensitivity information  $\delta\mathbf{W} = \alpha\nabla_{\mathbf{W}}e(\mathbf{x}^a)$  may not preserve this property, however, a perturbation  $\delta\mathbf{W}$  induced from a matrix decomposition can. To address this issue, the following three matrix decompositions are considered.

- $\mathbf{W} = \mathbf{W}^{1/2}\mathbf{W}^{T/2}$ , where  $\mathbf{W}^{1/2}$  is a square root of  $\mathbf{W}$ .
- $\mathbf{W} = \mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}$ , where  $\mathbf{\Sigma} = \text{diag}(\boldsymbol{\sigma}^w)$  and  $\mathbf{C}$  are the standard deviation and correlation matrices, respectively.
- $\mathbf{W} = (\mathbf{\Sigma}\mathbf{C}^{1/2})(\mathbf{\Sigma}\mathbf{C}^{1/2})^T$ , where  $\mathbf{C} = \mathbf{C}^{1/2}\mathbf{C}^{T/2}$  so that  $\mathbf{W}^{1/2} = \mathbf{\Sigma}\mathbf{C}^{1/2}$ .

In these decompositions, a perturbation in  $\mathbf{W}^{1/2}$ ,  $\mathbf{\Sigma}$ , or  $\mathbf{C}^{1/2}$  will induce a perturbation in  $\mathbf{W}$  that is symmetric positive semi-definite, preserving the requisite structure of  $\mathbf{W}$ . The equations to evaluate the forecast sensitivities associated with these matrix decompositions are detailed next.

### 5.3.1 Sensitivity to the Covariance Square Root Specification

From the matrix decomposition  $\mathbf{W} = \mathbf{W}^{1/2}\mathbf{W}^{T/2}$ , a perturbation  $\delta\mathbf{W}^{1/2}$  is related to a perturbation  $\delta\mathbf{W}$  via

$$\delta\mathbf{W} = (\delta\mathbf{W}^{1/2})\mathbf{W}^{T/2} + \mathbf{W}^{1/2}(\delta\mathbf{W}^{1/2})^T. \quad (5.22)$$

Equation (5.22) is used to get the forecast sensitivity to  $\mathbf{W}^{1/2}$ .

$$\langle \nabla_{\mathbf{W}} e(\mathbf{x}^a), \delta\mathbf{W} \rangle = \langle \mathbf{v}_1 \mathbf{v}_2^T, (\delta\mathbf{W}^{1/2})\mathbf{W}^{T/2} + \mathbf{W}^{1/2}(\delta\mathbf{W}^{1/2})^T \rangle \quad (5.23)$$

Using the properties of the inner product (5.7), (5.23) can be written as

$$\langle \mathbf{v}_1 \mathbf{v}_2^T, (\delta\mathbf{W}^{1/2})\mathbf{W}^{T/2} + \mathbf{W}^{1/2}(\delta\mathbf{W}^{1/2})^T \rangle = \langle (\mathbf{v}_1 \mathbf{v}_2^T + \mathbf{v}_2 \mathbf{v}_1^T) \mathbf{W}^{1/2}, \delta\mathbf{W}^{1/2} \rangle, \quad (5.24)$$

which provides the forecast sensitivity to  $\mathbf{W}^{1/2}$ .

$$\nabla_{\mathbf{W}^{1/2}} e(\mathbf{x}^a) = (\mathbf{v}_1 \mathbf{v}_2^T + \mathbf{v}_2 \mathbf{v}_1^T) \mathbf{W}^{1/2} \quad (5.25)$$

Notice that (5.25) is a rank two matrix.

### 5.3.2 Sensitivity to the Standard Deviation Specification

A perturbation  $\delta\mathbf{\Sigma}$  induces a perturbation

$$\delta\mathbf{W}^{1/2} = (\delta\mathbf{\Sigma})\mathbf{C}^{1/2}. \quad (5.26)$$

The first-order variation is

$$\delta e = \langle (\mathbf{v}_1\mathbf{v}_2^T + \mathbf{v}_2\mathbf{v}_1^T)\mathbf{W}^{1/2}, \delta\mathbf{W}^{1/2} \rangle = \langle (\mathbf{v}_1\mathbf{v}_2^T + \mathbf{v}_2\mathbf{v}_1^T), (\delta\mathbf{\Sigma})\mathbf{C}^{1/2} \rangle. \quad (5.27)$$

From (5.7), (5.27) is simplified by using the properties of the trace operator.

$$\begin{aligned} \delta e &= \text{Tr} [(\mathbf{v}_1\mathbf{v}_2^T + \mathbf{v}_2\mathbf{v}_1^T)\mathbf{W}^{1/2}\mathbf{C}^{T/2}(\delta\mathbf{\Sigma})] \\ &= \text{Tr} [\delta\mathbf{\Sigma} (\mathbf{v}_2(\mathbf{C}^{1/2}\mathbf{W}^{T/2}\mathbf{v}_1)^T + \mathbf{v}_1(\mathbf{C}^{1/2}\mathbf{W}^{T/2}\mathbf{v}_2)^T)] \end{aligned} \quad (5.28)$$

Since  $\mathbf{\Sigma}$  is a diagonal matrix,  $\delta\mathbf{\Sigma}$  relates to a perturbation  $\delta\boldsymbol{\sigma}^w$  so that

$$\delta e = \text{Tr} [(\delta\boldsymbol{\sigma}^w \circ \mathbf{v}_2)(\mathbf{C}^{1/2}\mathbf{W}^{T/2}\mathbf{v}_1)^T] + \text{Tr} [(\delta\boldsymbol{\sigma}^w \circ \mathbf{v}_1)(\mathbf{C}^{1/2}\mathbf{W}^{T/2}\mathbf{v}_2)^T]. \quad (5.29)$$

The property that  $\text{Tr}(\mathbf{a}\mathbf{b}^T) = \mathbf{b}^T\mathbf{a}$  further simplifies (5.29), in which case

$$\delta e = (\mathbf{C}^{1/2}\mathbf{W}^{T/2}\mathbf{v}_1)^T(\delta\boldsymbol{\sigma}^w \circ \mathbf{v}_2) + (\mathbf{C}^{1/2}\mathbf{W}^{T/2}\mathbf{v}_2)^T(\delta\boldsymbol{\sigma}^w \circ \mathbf{v}_1), \quad (5.30)$$

where  $\circ$  denotes the Hadamard product of matrices. After additional simplification,

$$\delta e = \langle \mathbf{v}_2 \circ (\mathbf{C}\mathbf{\Sigma}\mathbf{v}_1) + \mathbf{v}_1 \circ (\mathbf{C}\mathbf{\Sigma}\mathbf{v}_2), \delta\boldsymbol{\sigma}^w \rangle, \quad (5.31)$$

which gives the forecast sensitivity to the standard deviation  $\boldsymbol{\sigma}^w$ .

$$\nabla_{\boldsymbol{\sigma}^w} e(\mathbf{x}^a) = \mathbf{v}_2 \circ (\mathbf{C}\boldsymbol{\Sigma}\mathbf{v}_1) + \mathbf{v}_1 \circ (\mathbf{C}\boldsymbol{\Sigma}\mathbf{v}_2) \quad (5.32)$$

### 5.3.3 Sensitivity to the Correlation Specification

A perturbation in the correlation matrix  $\mathbf{C}$  induces a perturbation of  $\mathbf{W}$  via

$$\delta\mathbf{W} = \boldsymbol{\Sigma}(\delta\mathbf{C})\boldsymbol{\Sigma}. \quad (5.33)$$

Once again, the properties of the inner product (5.7) imply that

$$\langle \nabla_{\mathbf{W}} e(\mathbf{x}^a), \delta\mathbf{W} \rangle = \langle \mathbf{v}_1 \mathbf{v}_2^T, \boldsymbol{\Sigma}(\delta\mathbf{C})\boldsymbol{\Sigma} \rangle = \langle \boldsymbol{\Sigma}\mathbf{v}_1 \mathbf{v}_2^T \boldsymbol{\Sigma}, \delta\mathbf{C} \rangle. \quad (5.34)$$

Therefore, the forecast sensitivity to the correlation matrix  $\mathbf{C}$  is

$$\nabla_{\mathbf{C}} e(\mathbf{x}^a) = \boldsymbol{\Sigma}\mathbf{v}_1 \mathbf{v}_2^T \boldsymbol{\Sigma} = (\boldsymbol{\Sigma}\mathbf{v}_1)(\boldsymbol{\Sigma}\mathbf{v}_2)^T. \quad (5.35)$$

Like the  $\mathbf{W}$ -sensitivity equation, the  $\mathbf{C}$ -sensitivity equation has rank one. Further,  $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\sigma}^w)$ , so the  $\mathbf{C}$ -sensitivity equation can be expressed using  $\boldsymbol{\sigma}^w$  as

$$\nabla_{\mathbf{C}} e(\mathbf{x}^a) = (\boldsymbol{\sigma}^w \circ \mathbf{v}_1)(\boldsymbol{\sigma}^w \circ \mathbf{v}_2)^T. \quad (5.36)$$

### 5.3.4 Sensitivity to the Correlation Square Root Specification

The sensitivity of the correlation model square root  $\mathbf{C}^{1/2}$  can also be obtained. A perturbation in  $\mathbf{C}^{1/2}$  induces a perturbation in  $\mathbf{W}^{1/2}$  via

$$\delta\mathbf{W}^{1/2} = \boldsymbol{\Sigma}(\delta\mathbf{C}^{1/2}). \quad (5.37)$$

Then

$$\delta e = \langle (\mathbf{v}_1 \mathbf{v}_2^T + \mathbf{v}_2 \mathbf{v}_1^T) \mathbf{W}^{1/2}, \boldsymbol{\Sigma}(\delta \mathbf{C}^{1/2}) \rangle = \langle \boldsymbol{\Sigma}(\mathbf{v}_1 \mathbf{v}_2^T + \mathbf{v}_2 \mathbf{v}_1^T) \mathbf{W}^{1/2}, \delta \mathbf{C}^{1/2} \rangle \quad (5.38)$$

implies that the forecast sensitivity to  $\mathbf{C}^{1/2}$  is

$$\nabla_{\mathbf{C}^{1/2}} e(\mathbf{x}^a) = \boldsymbol{\Sigma}(\mathbf{v}_1 \mathbf{v}_2^T + \mathbf{v}_2 \mathbf{v}_1^T) \mathbf{W}^{1/2}. \quad (5.39)$$

### 5.3.5 Summary of Sensitivity Equations

The forecast sensitivities to the model error parameters can be extracted from the block structures of  $\mathbf{z}$ ,  $\mathbf{W}$ , and the sensitivity equations. For the model error parameters at time  $t_i$ ,

$$\mathbf{v}_1 = \mathbf{Q}_i^{-1}[\boldsymbol{\chi}_i - \mathbf{M}_i^a \boldsymbol{\chi}_{i-1}], \quad (5.40a)$$

$$\mathbf{v}_2 = \mathbf{Q}_i^{-1}[\mathbf{x}_i^a - \mathcal{M}_i(\mathbf{x}_{i-1}^a) - \mathbf{q}_i]. \quad (5.40b)$$

Then, the sensitivity equations for the model error parameters, summarized in Table 5.1, follow from equations (5.17), (5.19), (5.25), (5.32), (5.36), and (5.39).

## 5.4 Sensitivities Using a Single Outer Loop Iteration

The model error sensitivity equations can be simplified in the case when the analysis is obtained from a single outer loop iteration of the incremental algorithm. Since the quadratic approximation  $\hat{J}(\mathbf{x})$  of (5.2) is used instead, obtained by linearizing the model forecast and observation operators about a sequence of guess states



Parameter	Significance	Dimension	Forecast Sensitivity
$\mathbf{q}$	Model error bias	$\mathbb{R}^n$	$\mathbf{v}_1$
$\boldsymbol{\sigma}^q$	Model error standard deviation	$\mathbb{R}^n$	$\mathbf{v}_2 \circ (\mathbf{C}\boldsymbol{\Sigma}\mathbf{v}_1) + \mathbf{v}_1 \circ (\mathbf{C}\boldsymbol{\Sigma}\mathbf{v}_2)$
$\mathbf{Q}$	Model error covariance	$\mathbb{R}^{n \times n}$	$\mathbf{v}_1 \mathbf{v}_2^T$
$\mathbf{Q}^{1/2}$	Model error covariance square root	$\mathbb{R}^{n \times n}$	$(\mathbf{v}_1 \mathbf{v}_2^T + \mathbf{v}_2 \mathbf{v}_1^T) \mathbf{Q}^{1/2}$
$\mathbf{C}$	Model error correlation	$\mathbb{R}^{n \times n}$	$(\boldsymbol{\sigma}^q \circ \mathbf{v}_1)(\boldsymbol{\sigma}^q \circ \mathbf{v}_2)^T$
$\mathbf{C}^{1/2}$	Model error correlation square root	$\mathbb{R}^{n \times n}$	$\boldsymbol{\Sigma} (\mathbf{v}_1 \mathbf{v}_2^T + \mathbf{v}_2 \mathbf{v}_1^T) \mathbf{Q}^{1/2}$

Table 5.1: Summary of forecast error sensitivities to model error input parameters in w4D-Var.

$\{\mathbf{x}_0^g, \dots, \mathbf{x}_N^g\}$ ,  $\boldsymbol{\Gamma}(\mathbf{x})$  is replaced by the first-order linearization about  $\mathbf{x}^g$

$$\boldsymbol{\Gamma}(\mathbf{x}) \approx \boldsymbol{\Gamma}(\mathbf{x}^g) + \boldsymbol{\Gamma}_{\mathbf{x}}(\mathbf{x}^g)(\mathbf{x} - \mathbf{x}^g). \quad (5.41)$$

Then, the first-order optimality condition is  $\nabla_{\mathbf{x}} \hat{J}(\mathbf{x}^a) = \mathbf{0}$ , where

$$\hat{J}(\mathbf{x}) = \frac{1}{2} [\boldsymbol{\Gamma}(\mathbf{x}^g) + \boldsymbol{\Gamma}_{\mathbf{x}}(\mathbf{x}^g)(\mathbf{x} - \mathbf{x}^g) - \mathbf{z}]^T \mathbf{W}^{-1} [\boldsymbol{\Gamma}(\mathbf{x}^g) + \boldsymbol{\Gamma}_{\mathbf{x}}(\mathbf{x}^g)(\mathbf{x} - \mathbf{x}^g) - \mathbf{z}]. \quad (5.42)$$

In this framework, the sensitivity equations in Table 5.1 remain valid if the following changes are made.

- The linear system (5.15) is replaced by

$$\left[ \nabla_{\mathbf{xx}}^2 \hat{J}(\mathbf{x}^a) \right] \hat{\boldsymbol{\chi}} = \nabla_{\mathbf{x}} e(\mathbf{x}^a), \quad (5.43)$$

where

$$\nabla_{\mathbf{x}\mathbf{x}}^2 \hat{J}(\mathbf{x}^a) = \mathbf{\Gamma}_{\mathbf{x}}^T(\mathbf{x}^g) \mathbf{W}^{-1} \mathbf{\Gamma}_{\mathbf{x}}(\mathbf{x}^g) \quad (5.44)$$

is the Hessian of (5.42), and  $\boldsymbol{\chi}$  is replaced by  $\hat{\boldsymbol{\chi}}$ . The Hessian of the quadratic approximation  $\nabla_{\mathbf{x}\mathbf{x}}^2 \hat{J}(\mathbf{x}^a)$  consists only of first-order tangent linear and adjoint models of the forecast and observation models such that higher-order derivatives and the development of a second-order adjoint model are not required to evaluate  $\hat{\boldsymbol{\chi}}$ .

- The vectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$  from (5.40) are replaced by

$$\hat{\mathbf{v}}_1 = \mathbf{Q}_i^{-1} [\hat{\boldsymbol{\chi}}_i - \mathbf{M}_i^g \hat{\boldsymbol{\chi}}_{i-1}] \quad (5.45a)$$

$$\hat{\mathbf{v}}_2 = \mathbf{Q}_i^{-1} [\mathbf{x}_i^a - \mathcal{M}_i(\mathbf{x}_{i-1}^g) - \mathbf{M}_i^g(\mathbf{x}_{i-1}^a - \mathbf{x}_{i-1}^g) - \mathbf{q}_i]. \quad (5.45b)$$

## 5.5 Adaptive Tuning of Model Error Parameters

An adjoint-based approach has been considered by Song et al. [73] to achieve adaptive tuning of the background error covariance model in a hybrid ensemble Kalman filter/4D-Var DAS. The forecast error sensitivity equations provide a tool for tuning the model error covariance matrix specification. The forecast error gradients to  $\mathbf{Q}^{1/2}$ ,  $\boldsymbol{\sigma}^q$ , and  $\mathbf{C}^{1/2}$  give information on reducing the forecast error using the steepest descent direction in an update of the form

$$\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} - \alpha_k \nabla_{\mathbf{X}} e(\mathbf{x}^a) \Big|_{\mathbf{X}=\mathbf{X}^{(k)}} \quad (5.46)$$

and the specification of an updated model error covariance  $\mathbf{Q}^{(k+1)}$  based on the new value of the parameter  $\mathbf{Q}^{1/2}$ ,  $\boldsymbol{\sigma}^q$ , or  $\mathbf{C}^{1/2}$ . A practical approach for adaptive tuning of

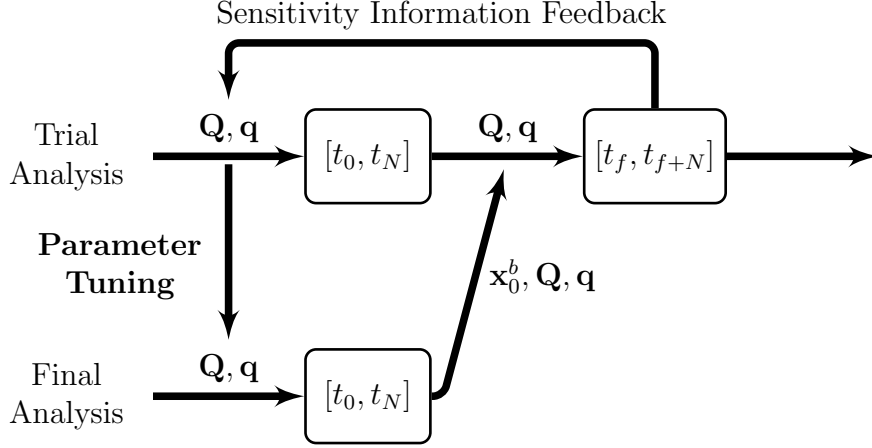


Figure 5.1: A visual of the model error parameter tuning procedure.

model error parameters and the selection of the step length  $\alpha_k$  are presented below.

After assimilating data over the window  $[t_0, t_N]$ , the analysis sequence for the next window  $[t_{N+1}, t_{2N+1}]$  will be used as the verification state  $\mathbf{x}^v$  in the forecast error functional, which is needed in the evaluation of the error sensitivity equations. This information is then fed back to the first window  $[t_0, t_N]$  to produce a new model error covariance specification and a new analysis sequence is obtained by re-assimilating the same observational data. The quality of the new specification is evaluated by comparing the errors in the forecasts initiated from the analyses produced in each of the two assimilation systems. If a sufficient reduction in the forecast error is achieved, the new DAS is accepted. The evaluation of the sensitivities and the quality of the updated DAS is computationally expensive, so it is impractical to implement more than one iteration of the gradient descent algorithm per assimilation cycle. This means that if there was not a sufficient decrease in the forecast error, the new DAS is rejected. Figure 5.1 shows a diagram of the sensitivity feedback loop.

The step-size for the gradient descent is initialized to  $\alpha_0 = 1$  and is dynamically updated between cycles. After a successful reduction in the forecast error, the step-

size is updated according to [68, p. 59]

$$\alpha = \frac{2(e(\mathbf{x}^a) - e(\mathbf{x}_{\text{new}}^a))}{\|\nabla_{\mathbf{x}} e(\mathbf{x}^a)\|^2} \quad (5.47)$$

which is the step-size obtained as the minimizer of the quadratic interpolation using the data  $e(\mathbf{x}^a)$ ,  $e(\mathbf{x}_{\text{new}}^a)$ , and  $\|\nabla_{\mathbf{x}} e(\mathbf{x}^a)\|^2$ . If the new DAS is rejected, the step-size is halved for the next cycle instead of being set by (5.47). Algorithm 5.1 summarizes this sensitivity procedure.

In general, the  $\mathbf{Q}^{1/2}$ - and  $\mathbf{C}^{1/2}$ -sensitivity matrices are full matrices and do not conform to any matrix structure. If  $\mathbf{Q}$  is a band matrix, then updating the DAS using these sensitivity matrices will not preserve this structure. This can be remedied through a localization operator, as used in the ensemble-based error covariance specification. Modifying the model error sensitivity equations in Table 5.1 to account for a band matrix structure simply requires setting the sensitivities outside the diagonal band to zero.

## 5.6 Numerical Results

Again, preliminary numerical results are presented for a DAS using the Lorenz models from in Section 1.4. For these experiments, a data assimilation window consists of the current time  $t_0$  and three 6-hour time-steps,  $[t_0, t_3]$ . Observational data  $\mathbf{y}_i$  at time  $t_i$  are taken at locations  $10i + k$  for  $k = 1, 2, \dots, 10$  so that  $\mathbf{y}_i$  is a 10-dimensional vector. Thus, the data  $\mathbf{y}_0$  are located at locations 1 through 10,  $\mathbf{y}_1$  are at locations 11 through 20, and so on. Through this rotating observing system, each grid point is observed during the assimilation cycle. All observations are generated from the true state with the observational error taken from the distribution  $N(0, (\sigma^o)^2)$  with the standard deviation specified as  $\sigma^o = 0.55$ . The background error covariance is defined

---

**Algorithm 5.1:** Procedure to update the model error covariance matrix using sensitivities.

---

```

1: procedure UPDATE QMODEL( $\mathbf{X}_0, \alpha_0$ )
2:    $\mathbf{Q} = \mathbf{Q}(\mathbf{X}_0)$  ▷ Initialize  $\mathbf{Q}$  using parameter  $\mathbf{X}_0$ 
3:    $\alpha = \alpha_0$  ▷ Set initial step-size
4:    $\mathbf{x}^a = \text{w4DVar}(\mathbf{x}_0^b, \mathbf{y}, \mathbf{B}, \mathbf{R}, \mathbf{Q}, \mathbf{q})$ 

5:    $\mathbf{p} = -\nabla_{\mathbf{x}} e(\mathbf{x}^a)$  ▷ Set the search direction
6:    $\mathbf{X}_{\text{new}} = \mathbf{X}_0 + \alpha \mathbf{p}$ 

7:    $\mathbf{Q}_{\text{new}} = \mathbf{Q}(\mathbf{X}_{\text{new}})$ 
8:    $\mathbf{x}_{\text{new}}^a = \text{w4DVar}(\mathbf{x}_0^b, \mathbf{y}, \mathbf{B}, \mathbf{R}, \mathbf{Q}_{\text{new}}, \mathbf{q})$  ▷ Redo analysis with new specification
9:    $\varepsilon = e(\mathbf{x}_{\text{new}}^a) - e(\mathbf{x}^a)$  ▷ Compute forecast error impact

10:  if  $\varepsilon < 0$  then ▷ Forecast error was decreased
11:     $\mathbf{Q} = \mathbf{Q}_{\text{new}}$  ▷ Update model error covariance specification
12:     $\mathbf{x}^a = \mathbf{x}_{\text{new}}^a$  ▷ Update analysis state
13:     $\alpha = 2(e(\mathbf{x}^a) - e(\mathbf{x}_{\text{new}}^a)) / \|\nabla_{\mathbf{x}} e(\mathbf{x}^a)\|^2$  ▷ Update step-size for the next cycle
14:  else ▷ Reject the updated specification
15:     $\alpha = 0.5\alpha$  ▷ Reduce step-size for next cycle
16:  end if

17:  Continue to next assimilation cycle
18: end procedure

```

---

to be a static band matrix with correlation structure specified by (3.9) with correlation parameter  $c = 4$ . The background error variance is set to be  $(\sigma^b)^2 = 0.1$  for each grid point. Initially, the background  $\mathbf{x}_0^b$  is the truth perturbed by random noise and is a forecast of a previous analysis for future cycles. A visual of the rotating observing system and the background error covariance specification are shown in Figure 5.2.

### 5.6.1 Numerical Convergence Tests

The validity of the model error sensitivity equations is established using a numerical gradient check. For this test, the forecast error aspect is defined similar to (5.1)

$$e(\mathbf{x}_0^a, \dots, \mathbf{x}_N^a) = (\mathbf{x}_N^f - \mathbf{x}_{N+4}^v)^T \mathbf{E} (\mathbf{x}_N^f - \mathbf{x}_{N+4}^v), \quad (5.48)$$

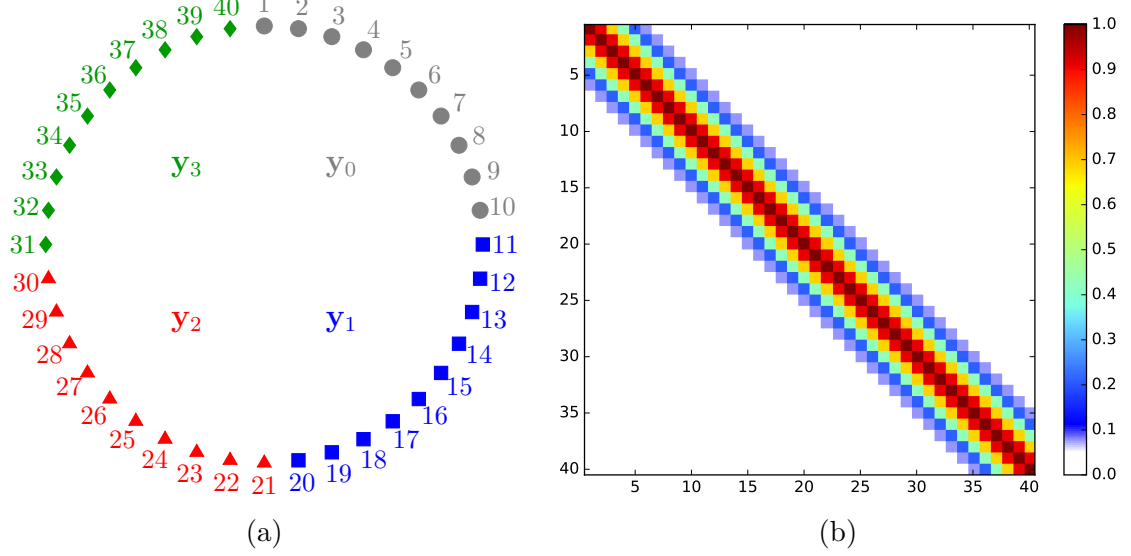


Figure 5.2: (a) A visual of the rotating observing system, (b) The specification of the background error correlation matrix.

where  $\mathbf{E} = \mathbf{I}_n$ ,  $\mathbf{x}_N^f = \mathcal{M}_{N \rightarrow N+4}(\mathbf{x}_N^a)$  is the 24-hour forecast of  $\mathbf{x}_N^a$ , and  $\mathbf{x}_{N+4}^v$  is the verifying analysis at time  $t_{N+4}$ . The forecast sensitivity of (5.48) to the analysis is

$$\nabla_{\mathbf{x}} e(\mathbf{x}^a) = \left[ \mathbf{0}^T \quad \cdots \quad \mathbf{0}^T \quad [2\mathbf{M}_{N \rightarrow N+4}^T \mathbf{E}(\mathbf{x}_N^f - \mathbf{x}_{N+4}^v)]^T \right]^T \quad (5.49)$$

where  $\mathbf{M}_{N \rightarrow N+4}^T$  is the adjoint of the tangent linear forecast model from time  $t_{N+4}$  to  $t_N$  evaluated at  $\mathbf{x}_N^a$ .

The convergence of the second-order accurate finite difference approximation

$$\langle \nabla_{\mathbf{x}} e(\mathbf{X}), \Delta \mathbf{X} \rangle = \frac{e(\mathbf{X} + h\Delta \mathbf{X}) - e(\mathbf{X} - h\Delta \mathbf{X})}{2h} + \mathcal{O}(h^2) \quad (5.50)$$

as  $h \rightarrow 0$  provides the means to verify the model error parameter sensitivity equations derived in Section 5.1 by observing the behavior of the error in the finite difference

Perturbation	$\mathbf{Q}_{+,-}$
$h\Delta\mathbf{Q}$	$\mathbf{Q} \pm h\Delta\mathbf{Q}$
$h\Delta\mathbf{Q}^{1/2}$	$(\mathbf{Q}^{1/2} \pm h\Delta\mathbf{Q}^{1/2})(\mathbf{Q}^{1/2} \pm h\Delta\mathbf{Q}^{1/2})^T$
$h\Delta\mathbf{C}$	$\mathbf{Q} \pm h\Sigma\Delta\mathbf{C}\Sigma$
$h\Delta\mathbf{C}^{1/2}$	$\Sigma(\mathbf{C}^{1/2} \pm h\Delta\mathbf{C}^{1/2})(\mathbf{C}^{1/2} \pm h\Delta\mathbf{C}^{1/2})^T\Sigma$
$h\Delta\boldsymbol{\sigma}^q$	$(\Sigma \pm h\Delta\Sigma)\mathbf{C}(\Sigma \pm h\Delta\Sigma)$

Table 5.2: Specifications of  $\mathbf{Q}_+$  and  $\mathbf{Q}_-$  given a perturbation in  $\mathbf{Q}$  or one of its matrix factors.

approximation.

$$E_h = \frac{e(\mathbf{X} + h\Delta\mathbf{X}) - e(\mathbf{X} - h\Delta\mathbf{X})}{2h} - \langle \nabla_{\mathbf{X}} e(\mathbf{X}), \Delta\mathbf{X} \rangle \quad (5.51)$$

For  $h = 2^{-n}$ , the logarithm of the error ratio  $\log_2(E_h/E_{h/2})$  should be approximately 2 as  $n$  increases for each of the model error sensitivity equations. The forecast sensitivities are evaluated for a DAS where the model error covariance matrices are prescribed as  $\mathbf{Q}_i = \text{diag}(\mathbf{Q}^t)$ . The numerical experiment for this gradient check is implemented as follows.

- Obtain the analysis sequence and the forecast sensitivity.
- Setup two new systems where the model error covariances are prescribed by

$$\mathbf{Q}_+ = \mathbf{Q}(\mathbf{X} + h\Delta\mathbf{X}), \quad \mathbf{Q}_- = \mathbf{Q}(\mathbf{X} - h\Delta\mathbf{X}) \quad (5.52)$$

where  $\mathbf{X}$  is a stand-in for the model error parameter being perturbed. Table 5.2 summarizes these specifications for each model error covariance parameter.

When perturbing  $\mathbf{q}$ , prescribe

$$\mathbf{q}_+ = \mathbf{q} + h\Delta\mathbf{q}, \quad \mathbf{q}_- = \mathbf{q} - h\Delta\mathbf{q} \quad (5.53)$$

instead.

- Obtain the analysis sequences for the new systems and evaluate the forecast errors.
- Form the table of finite difference error ratios.

Table 5.3 shows the finite difference errors for the model error parameter sensitivity equations. The logarithm of the error ratios approaches 2 for each of the sensitivity equations, agreeing with the correct rate of convergence of (5.50). This provides a numerical validation of the sensitivity equations. The range of  $h$  is different for the  $\mathbf{Q}$ -sensitivity experiment since for larger values of  $h$ ,  $\mathbf{Q}_-$  was not positive definite and no analyses have been produced in those cases.

### 5.6.2 Adaptive Tuning of the Model Error Covariance Specification

A comparative analysis is done to investigate the performance of a w4D-Var DAS when the specification of the model error covariance is dynamically updated using the tuning procedure from Section 5.5. Two systems are considered where the model error covariance matrices are prescribed as follows.

- (Control)  $\mathbf{Q} = \text{diag}(\mathbf{Q}^t)$  and will remain static throughout the experiment.
- (Sensitivity)  $\mathbf{Q} = \text{diag}(\mathbf{Q}^t)$  for the first assimilation cycle and is updated using the gradient descent algorithm via sensitivities to the square root factor  $\mathbf{Q}^{1/2}$ .



$h$	$E_h$	$\log_2(E_h/E_{h/2})$		$h$	$E_h$	$\log_2(E_h/E_{h/2})$	
$2^{-14}$	$4.98 \times 10^{-3}$	—		$2^{-7}$	$3.93 \times 10^{-5}$	—	
$2^{-15}$	$1.24 \times 10^{-3}$	2.00		$2^{-8}$	$9.82 \times 10^{-6}$	2.00	
$2^{-16}$	$3.11 \times 10^{-4}$	2.00		$2^{-9}$	$2.46 \times 10^{-6}$	2.00	
$2^{-17}$	$7.78 \times 10^{-5}$	2.00		$2^{-10}$	$6.14 \times 10^{-7}$	2.00	
$2^{-18}$	$1.94 \times 10^{-5}$	2.00		$2^{-11}$	$1.53 \times 10^{-7}$	2.00	
(a) $\mathbf{Q}$ -sensitivity				(b) $\mathbf{C}$ -sensitivity			
$h$	$E_h$	$\log_2(E_h/E_{h/2})$		$h$	$E_h$	$\log_2(E_h/E_{h/2})$	
$2^{-9}$	$3.49 \times 10^{-1}$	—		$2^{-5}$	$6.19 \times 10^{-2}$	—	
$2^{-10}$	$8.73 \times 10^{-2}$	2.00		$2^{-6}$	$1.56 \times 10^{-2}$	1.99	
$2^{-11}$	$2.18 \times 10^{-2}$	2.00		$2^{-7}$	$3.89 \times 10^{-3}$	2.00	
$2^{-12}$	$5.46 \times 10^{-3}$	2.00		$2^{-8}$	$9.74 \times 10^{-4}$	2.00	
$2^{-13}$	$1.37 \times 10^{-3}$	2.00		$2^{-9}$	$2.43 \times 10^{-4}$	2.00	
(c) $\mathbf{Q}^{1/2}$ -sensitivity				(d) $\mathbf{C}^{1/2}$ -sensitivity			
$h$	$E_h$	$\log_2(E_h/E_{h/2})$		$h$	$E_h$	$\log_2(E_h/E_{h/2})$	
$2^{-5}$	$2.08 \times 10^0$	—		$2^{-5}$	$9.02 \times 10^{-5}$	—	
$2^{-6}$	$5.26 \times 10^{-1}$	1.99		$2^{-6}$	$2.25 \times 10^{-5}$	2.00	
$2^{-7}$	$1.32 \times 10^{-1}$	2.00		$2^{-7}$	$5.63 \times 10^{-6}$	2.00	
$2^{-8}$	$3.30 \times 10^{-2}$	2.00		$2^{-8}$	$1.41 \times 10^{-6}$	2.00	
$2^{-9}$	$8.24 \times 10^{-3}$	2.00		$2^{-9}$	$3.52 \times 10^{-7}$	2.00	
(e) $\sigma^q$ -sensitivity				(f) $\mathbf{q}$ -sensitivity			

Table 5.3: Finite difference errors and log error ratios for each of the model error sensitivity equations.

The w4D-Var forecast error functional for this experiment is defined by

$$e(\mathbf{x}_0^a, \dots, \mathbf{x}_N^a) = (\mathbf{x}_N^f - \mathbf{x}_{N+1}^v)^T \mathbf{E} (\mathbf{x}_N^f - \mathbf{x}_{N+1}^v), \quad (5.54)$$

where  $\mathbf{E} = \mathbf{I}_n$  and  $\mathbf{x}_N^f = \mathcal{M}_{N+1}(\mathbf{x}_N^a)$  is the 6-hour forecast of  $\mathbf{x}_N^a$  from time  $t_N$  to  $t_{N+1}$ . It is sufficient to only forecast  $\mathbf{x}_N^a$  one time-step to obtain the forecast sensitivity to  $\mathbf{Q}_i$  and  $\mathbf{q}_i$  since each analysis state  $\mathbf{x}_i^a$  depends on the model error statistics over the

entire window  $[t_0, t_N]$ . The forecast sensitivity of (5.54) to the analysis is

$$\nabla_{\mathbf{x}} e(\mathbf{x}^a) = \begin{bmatrix} \mathbf{0}^T & \dots & \mathbf{0}^T & [2\mathbf{M}_{N+1}^T \mathbf{E}(\mathbf{x}_N^f - \mathbf{x}_{N+1}^v)]^T \end{bmatrix}^T \quad (5.55)$$

where  $\mathbf{M}_{N+1}^T$  is the adjoint of the tangent linear forecast model from time  $t_{N+1}$  to  $t_N$  evaluated at  $\mathbf{x}_N^a$ . The 6-hour forecast lead time is used for quasi real time adaptivity of the parameters of the **Q**-model. The 6-hour lead time has been previously considered in the work of Hotta [46] to achieve flow-dependent proactive quality control using ensemble-based estimation of the sensitivity to observations. In his work, Hotta has also found that the 6-hour ensemble forecast sensitivity to observations is, at least qualitatively, consistent with the results based on 24-forecast sensitivities.

Initially, the step-size for the gradient descent is set to 1 and dynamically updated throughout the experiment. Localization is applied so that the dynamic model error covariance matrices have bandwidth 20.

Figure 5.3 shows the ratio of the monthly (30-day) average background and analysis errors to the control analysis error. An error ratio less than 1 indicates better performance than the control and greater than 1 means a worse performance. Also shown is the error ratio for the control background error, which is about 1.1 for the entire two years, meaning that the control background error is about 10% worse than the control analysis error. The background and analysis errors for the sensitivity DAS are trending downward during months 1 through 7. This time acts as a spin-up period for the sensitivity-based system and proper analysis of its performance can be made once the downward trend in the background and analysis errors has finished. At the end of this period, the analysis error ratio is about 0.5, implying a 50% analysis error reduction has taken place. Furthermore, the background error ratio is less than 1, and in particular, the background obtained from the sensitivity DAS estimates the

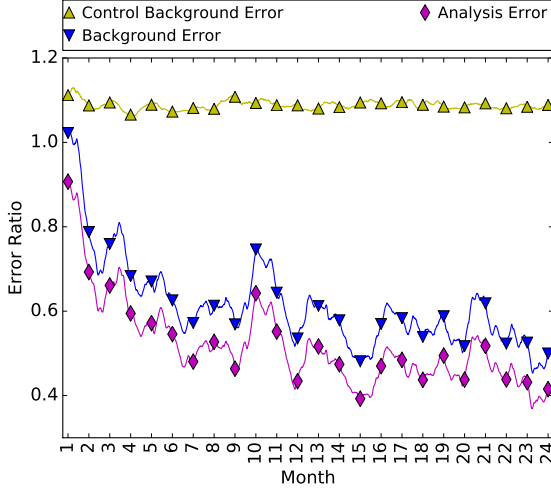


Figure 5.3: The ratio of the global monthly errors to the control w4D-Var experiment.

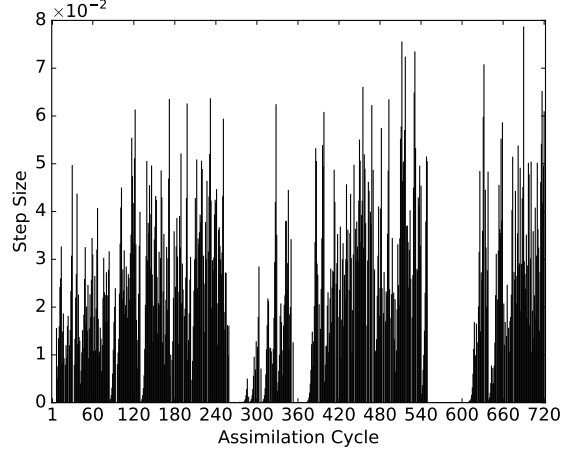


Figure 5.4: The step-size for the gradient descent update of the model error covariance matrices for each data assimilation cycle over a two-year period.

truth better than the analysis of the control. In total, the two-year period consisted of 720 assimilation cycles, for which there were 603 updates to  $\mathbf{Q}$  for a total of 83.75% success rate in reducing the forecast error. Figure 5.4 shows the step-sizes for each assimilation cycle. For the assimilation cycles in which the forecast error was not reduced, the step-size was recorded as  $\alpha = 0$ , signifying that  $\mathbf{Q}$  was not updated for the next assimilation cycle.

The sensitivity w4D-Var system performance is also compared to the ensemble-based w4D-Var approaches from Chapter 3. The model error statistics for the two additional assimilation systems are prescribed as follows.

- (Ensemble)  $\mathbf{Q}$  is specified as the flow-dependent estimate  $\mathbf{Q}_e$  for an ensemble size of  $N_e = 20$  with localization obtained from (3.9) with correlation parameter 4.
- (Hybrid)  $\mathbf{Q} = 0.5\mathbf{Q}_c + 0.5\mathbf{Q}_e$ , where the static component is specified as  $\mathbf{Q}_c = \text{diag}(\mathbf{Q}^t)$  and  $\mathbf{Q}_e$  is as specified for the weak ensemble DAS.

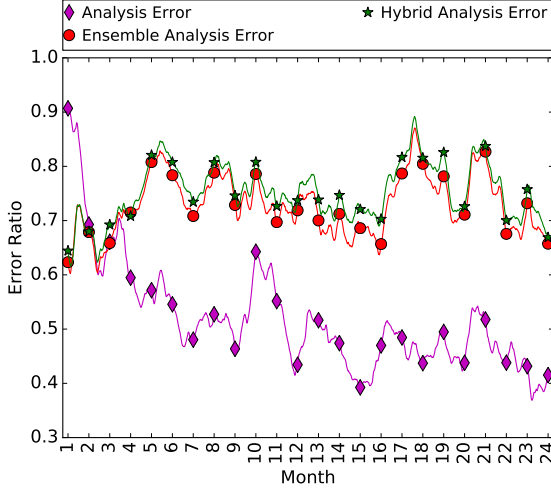


Figure 5.5: The same as Figure 5.3, but compares the monthly error ratios of the sensitivity DAS to the ensemble-based DASs.

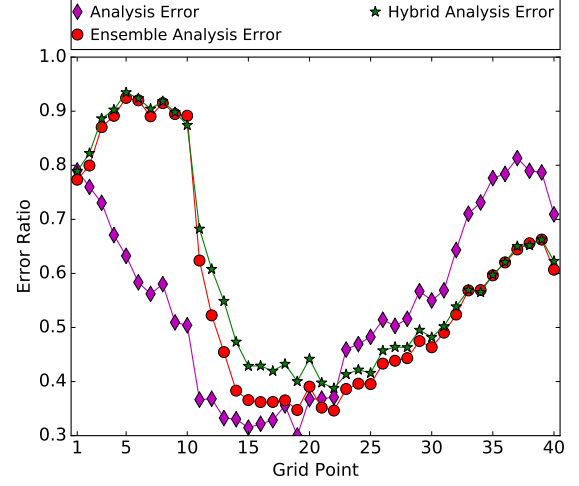


Figure 5.6: The ratio of the analysis errors for the sensitivity and ensemble-based systems over each grid point.

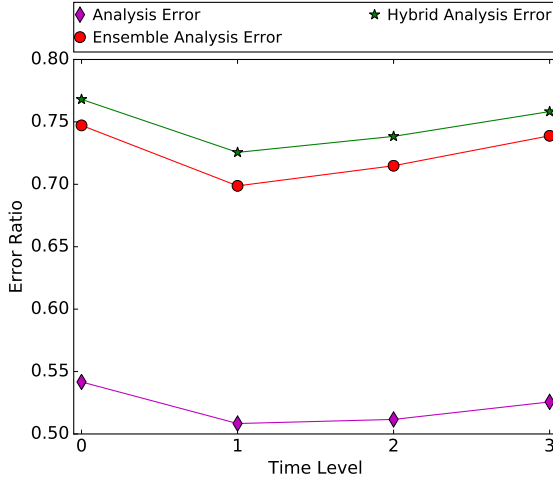


Figure 5.7: The ratio of the analysis errors to the control for each time in the data assimilation window.

data assimilation window  $[t_0, t_3]$  over two years. For the first three months, the ensemble-based systems outperform the sensitivity-based DAS.

The ensemble and hybrid w4D-Var schemes do not require a spin-up as they use an ensemble covariance matrix updated at every assimilation cycle. On the other hand,

The performance of the sensitivity w4D-Var and the ensemble-based w4D-Var systems is evaluated by comparing the analysis errors in three situations. Figure 5.5 shows the monthly analysis error ratios for each DAS to the control, like Figure 5.3, but does not include background errors. Figure 5.6 shows the analysis error ratios over each grid point while Figure 5.7 shows the average analysis error ratios for each time level in the

the sensitivity-driven update of  $\mathbf{Q}$  involves one low-rank matrix each cycle and only after a number of updates will a significant reduction in the analysis error be seen. For grid points 2 through 20, the sensitivity w4D-Var DAS has better performance than the ensemble-based schemes. The error ratios vary between about 0.3 in the middle of the domain to about 0.8 at the highest. The best performance is seen at grid points 11 through 20, which can be explained by the choice of observing system. In variational data assimilation, it is known that the analysis error over a data assimilation window  $[t_0, t_N]$  is at its lowest in the middle of the window, which is the second time level  $t_1$  in this experiment. This is corroborated by Figure 5.7 which plots the average analysis error ratios to the control for each time level in the data assimilation window. It is at time  $t_1$  in which grid points 11 through 20 are observed in the rotating observing system. Additionally, it is noted that Figure 5.7 indicates that the sensitivity w4D-Var has the best performance overall in  $[t_0, t_3]$ .

## 5.7 Conclusion

This chapter extends the theoretical framework of the adjoint-based sensitivity analysis in data assimilation to incorporate model error parameters in the weak-constraint 4D-Var formulation. Equations to evaluate the sensitivity of a forecast error aspect to the DAS representation of the model error bias, standard deviation, and correlation structures are derived from the first-order optimality condition associated with the nonlinear optimization problem. It is emphasized that an all-at-once evaluation of the sensitivities to various parameters in the  $\mathbf{Q}$ -model may be efficiently performed by properly exploiting the low-rank structure of the  $\mathbf{Q}$ -sensitivity matrix.

A new computationally feasible approach is introduced for tuning  $\mathbf{Q}$ -model parameters based on derivative information extracted from a short-range forecast error measure. Our novel approach to adaptive covariance tuning relies on a two-stage

procedure that includes a trial analysis followed by adjoint-based forecast error sensitivity and parameter tuning to obtain the final analysis. The information collected through a feedback mechanism is used to dynamically update the step-size of the steepest descent iteration. To the best of our knowledge, this is the first attempt to perform variable step-size gradient-based tuning of the model error covariance in a w4D-Var DAS. A proof-of-concept has been presented with the Lorenz model and, the simplicity of the model notwithstanding, results from a preliminary set of data assimilation experiments show the potential of this methodology to produce improved state estimates as compared with other approaches in data assimilation.

The theoretical superiority of the w4D-Var formulation can be achieved in practical applications only through a proper representation of the model error statistics. Therefore, it is expected that the techniques presented here will find an extended range of applications to assess and improve the performance of variational data assimilation and forecast systems.

## Chapter 6

### Experiments with the Shallow Water Equations

The second stage of numerical experiments takes place with the shallow water equations

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + \frac{\partial \phi}{\partial x} = f v - g \frac{\partial H}{\partial x} \quad (6.1a)$$

$$\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} = -f u \quad (6.1b)$$

$$\frac{\partial \phi}{\partial t} + u \frac{\partial \phi}{\partial x} + \phi \frac{\partial u}{\partial x} = 0 \quad (6.1c)$$

where  $x \in [0, L]$ . The variables  $u = u(x, t)$  and  $v = v(x, t)$  represent the eastward and northward components of velocity and  $\phi = \phi(x, t)$  represents the geopotential, defined by  $\phi = g\eta(x, t)$  with  $g$  being the acceleration due to gravity and  $\eta$  the depth of the fluid. The topography of the bottom is given by  $H(x)$ , and  $f$  is the Coriolis parameter. Setting the values of the parameters as  $f = 1.45842 \times 10^{-4} \sin \theta \text{ s}^{-1}$ ,  $g = 9.8 \text{ ms}^{-2}$ , and  $L = 40074 \cos \theta \text{ km}$  are suitable for a model approximating the atmosphere of Earth at  $\theta^\circ$  north latitude. Experiments in this chapter are performed for  $\theta = 30^\circ$ . This model has previously been used in the study of data assimilation techniques [41, 60]. See Figure 6.1 for a visual representation of the shallow water model.

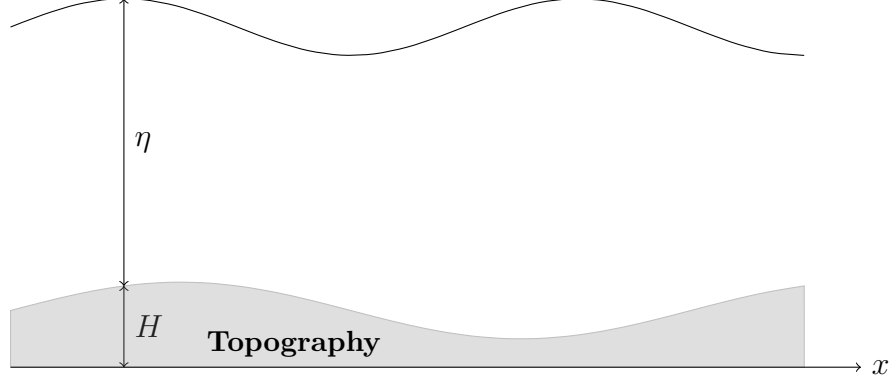


Figure 6.1: Schematic representation of the shallow water model.

## 6.1 The Discrete Model

For data assimilation experiments, the forecast model is obtained by discretizing the flux form of the shallow water equations using finite differences, which includes artificial diffusion with  $K = 500 \text{ m}^2\text{s}^{-1}$  to help eliminate spurious oscillations. The discrete model for the shallow water equations using forward time and centered spacing is given by

$$\begin{aligned}
 m_j^{k+1} = m_j^k - \frac{\Delta t}{4\Delta x} & [(u_{j+1}^k + u_j^k)(m_{j+1}^k + m_j^k) - (u_j^k + u_{j-1}^k)(m_j^k + m_{j-1}^k) + (\phi_{j+1}^k)^2 - \\
 & (\phi_{j-1}^k)^2 + g[(\phi_{j+1}^k + \phi_j^k)(H_{j+1} - H_j) + (\phi_j^k + \phi_{j-1}^k)(H_j - H_{j-1})]] + \Delta t f n_j^k \\
 & + K \frac{\Delta t}{\Delta x^2} [m_{j-1}^k - 2m_j^k + m_{j+1}^k], \tag{6.2a}
 \end{aligned}$$

$$\begin{aligned}
 n_j^{k+1} = n_j^k - \frac{\Delta t}{4\Delta x} & [(v_{j+1}^k + v_j^k)(m_{j+1}^k + m_j^k) - (v_j^k + v_{j-1}^k)(m_j^k + m_{j-1}^k)] - \Delta t f m_j^k \\
 & + K \frac{\Delta t}{\Delta x^2} [n_{j-1}^k - 2n_j^k + n_{j+1}^k], \tag{6.2b}
 \end{aligned}$$

$$\phi_j^{k+1} = \phi_j^k - \frac{\Delta t}{2\Delta x} [m_{j+1}^k - m_{j-1}^k] + K \frac{\Delta t}{\Delta x^2} [\phi_{j-1}^k - 2\phi_j^k + \phi_{j+1}^k], \tag{6.2c}$$

for  $j = 1, \dots, J$  spatial nodes so that the state vector  $\mathbf{x}_k$  has dimension  $n = 3J$ . In the discretized shallow water equations,  $m_j^k = \phi_j^k u_j^k$  and  $n_j^k = \phi_j^k v_j^k$ , where  $u_j^k \approx u(j\Delta x, k\Delta t)$ ,  $v_j^k \approx v(j\Delta x, k\Delta t)$ , and  $\phi_j^k \approx \phi(j\Delta x, k\Delta t)$ . The derivation of this



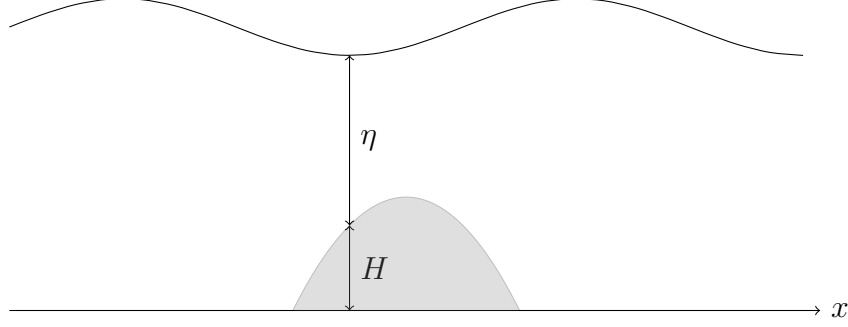


Figure 6.2: Shallow water diagram with bottom topography given by (6.3).

scheme is shown in Appendix B. Periodic spatial boundary conditions are imposed so that state  $J + 1$  corresponds to state 1 and so on. For the finite difference scheme, the spatial step-size is set to be  $\Delta x = L/J$  and the temporal step-size satisfies  $\Delta t = 0.1\Delta x$  to ensure numerical stability. The bottom topography  $H(x)$  consists of a ridge in the center of the domain defined, in meters, by

$$H(x) = \begin{cases} 0.5 \left[ 1 - \left( \frac{x - L/2}{10\Delta x} \right)^2 \right], & |x - L/2| \leq 10\Delta x \\ 0, & \text{otherwise} \end{cases} \quad (6.3)$$

as shown in Figure 6.2.

The true state for data assimilation experiments is represented by (6.2). By ignoring the bottom topography, the finite difference equations with  $H(x) = 0$  will only approximate the true state evolution and model error is introduced. Therefore,

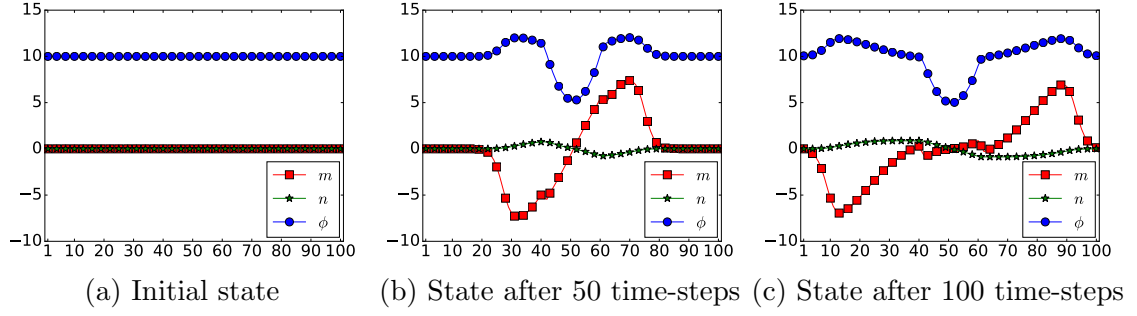


Figure 6.3: The “true” state of the shallow water model.

the forecast model  $\mathcal{M}_k$  is represented by

$$m_j^{k+1} = m_j^k - \frac{\Delta t}{4\Delta x} [(u_{j+1}^k + u_j^k)(m_{j+1}^k + m_j^k) - (u_j^k + u_{j-1}^k)(m_j^k + m_{j-1}^k) + (\phi_{j+1}^k)^2 - (\phi_{j-1}^k)^2] + \Delta t f n_j^k + K \frac{\Delta t}{\Delta x^2} [m_{j-1}^k - 2m_j^k + m_{j+1}^k], \quad (6.4a)$$

$$n_j^{k+1} = n_j^k - \frac{\Delta t}{4\Delta x} [(v_{j+1}^k + v_j^k)(m_{j+1}^k + m_j^k) - (v_j^k + v_{j-1}^k)(m_j^k + m_{j-1}^k)] - \Delta t f m_j^k + K \frac{\Delta t}{\Delta x^2} [n_{j-1}^k - 2n_j^k + n_{j+1}^k], \quad (6.4b)$$

$$\phi_j^{k+1} = \phi_j^k - \frac{\Delta t}{2\Delta x} [m_{j+1}^k - m_{j-1}^k] + K \frac{\Delta t}{\Delta x^2} [\phi_{j-1}^k - 2\phi_j^k + \phi_{j+1}^k]. \quad (6.4c)$$

The adjoint and tangent linear models of the forecast model (6.4) are shown in Appendix C.

## 6.2 Data Assimilation System

Numerical experiments will implement the methodologies of the previous chapters to gain insight on the model error statistics in this situation. The number of spatial nodes is chosen to be  $J = 100$  and the initial state of the system is given by

$$m_j^0 = 0 \text{ m}^3\text{s}^{-3}, \quad n_j^0 = 0 \text{ m}^3\text{s}^{-3}, \quad \phi_j^0 = 10 \text{ m}^2\text{s}^{-2}. \quad (6.5)$$

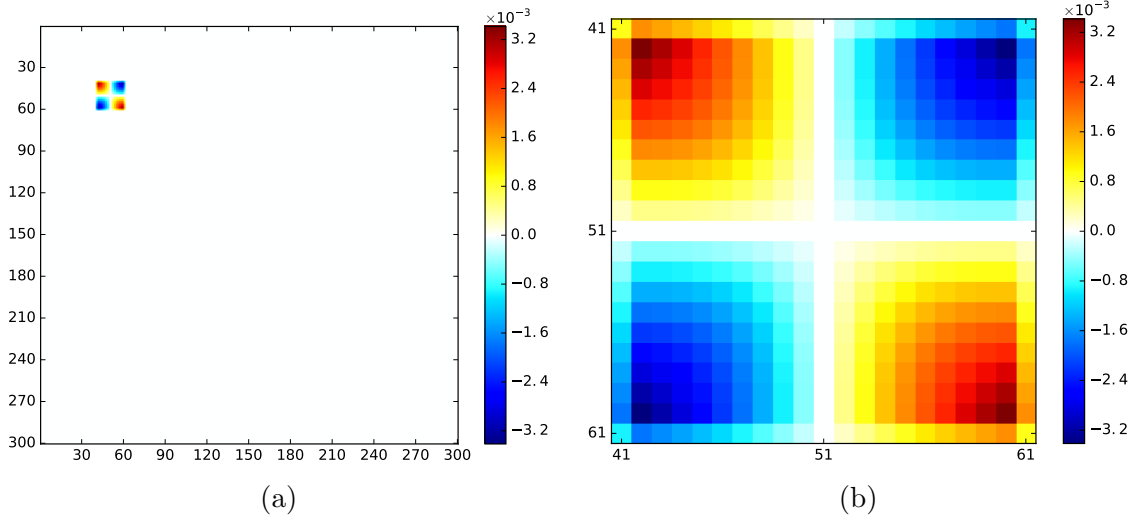


Figure 6.4: (a) The estimated model error covariance matrix for the discrete shallow water model, (b) Zoomed-in view of  $\mathbf{Q}^t$ .

The initial conditions correspond to a fixed point of the model equations (6.4) so the model state remains in equilibrium if data assimilation is not applied. Figure 6.3 shows the “true” state at different times, starting with the initial conditions (6.5). Figure 6.4 shows the model error covariance estimated using the sample covariance of model error realizations obtained from the difference between the truth (6.2) and the model (6.4). In these figures, the state vector  $\mathbf{x}_k$  is arranged so that states 1 through 100 correspond to  $m_{1:100}^k$ , states 101 through 200 correspond to  $n_{1:100}^k$ , and the final 100 correspond to  $\phi_{1:100}^k$ . Since the absence of bottom topography only affects the state of the  $m$ -variables, no model error exists in the forecast model for the  $n$ - and  $\phi$ -variables. In particular, the definition of the true bottom topography (6.3) implies that only  $m_{41:61}^k$  are affected by model error.

The DAS is defined such that a data assimilation window consists of four time levels  $[t_0, t_3]$  and observations are taken from the true state at every grid point and at every time level perturbed by unbiased random noise with covariance  $\mathbf{R}_k = \mathbf{I}_n$ . The background error and model error covariances are also specified as the identity

$h$	$E_h$	$\log_2(E_h/E_{h/2})$	$h$	$E_h$	$\log_2(E_h/E_{h/2})$
$2^{-9}$	$8.72 \times 10^{-2}$	—	$2^{-5}$	$2.97 \times 10^{-1}$	—
$2^{-10}$	$2.17 \times 10^{-2}$	2.00	$2^{-6}$	$7.37 \times 10^{-2}$	2.01
$2^{-11}$	$5.42 \times 10^{-3}$	2.00	$2^{-7}$	$1.83 \times 10^{-2}$	2.00
$2^{-12}$	$1.35 \times 10^{-3}$	2.00	$2^{-8}$	$4.59 \times 10^{-3}$	2.00
$2^{-13}$	$3.38 \times 10^{-4}$	2.00	$2^{-9}$	$1.14 \times 10^{-4}$	2.00
(a) $\mathbf{Q}$ -sensitivity			(b) $\mathbf{C}$ -sensitivity		
$h$	$E_h$	$\log_2(E_h/E_{h/2})$	$h$	$E_h$	$\log_2(E_h/E_{h/2})$
$2^{-9}$	$5.26 \times 10^{-2}$	—	$2^{-5}$	$1.51 \times 10^{-1}$	—
$2^{-10}$	$1.31 \times 10^{-2}$	1.99	$2^{-6}$	$3.78 \times 10^{-2}$	1.99
$2^{-11}$	$3.29 \times 10^{-3}$	1.99	$2^{-7}$	$9.46 \times 10^{-3}$	2.00
$2^{-12}$	$8.24 \times 10^{-4}$	2.00	$2^{-8}$	$2.36 \times 10^{-3}$	2.00
$2^{-13}$	$2.06 \times 10^{-4}$	2.00	$2^{-9}$	$5.91 \times 10^{-4}$	2.00
(c) $\mathbf{Q}^{1/2}$ -sensitivity			(d) $\mathbf{C}^{1/2}$ -sensitivity		
$h$	$E_h$	$\log_2(E_h/E_{h/2})$	$h$	$E_h$	$\log_2(E_h/E_{h/2})$
$2^{-5}$	$4.41 \times 10^{-2}$	—	$2^{-5}$	$7.71 \times 10^{-8}$	—
$2^{-6}$	$1.10 \times 10^{-2}$	2.00	$2^{-6}$	$1.92 \times 10^{-8}$	2.00
$2^{-7}$	$2.75 \times 10^{-3}$	2.00	$2^{-7}$	$4.81 \times 10^{-9}$	2.00
$2^{-8}$	$6.88 \times 10^{-4}$	2.00	$2^{-8}$	$1.20 \times 10^{-9}$	2.00
$2^{-9}$	$1.72 \times 10^{-4}$	2.00	$2^{-9}$	$2.95 \times 10^{-10}$	2.00
(e) $\sigma^q$ -sensitivity			(f) $\mathbf{q}$ -sensitivity		

Table 6.1: Finite difference errors and log error ratios for each of the model error sensitivity equations for the discrete shallow water model.

matrix and the initial background  $\mathbf{x}_0^b$  is the true state perturbed by random noise. The analysis is obtained using state space preconditioning (2.46), where the linear system is solved by the conjugate gradient method.

### 6.3 Numerical Results

The data assimilation approaches described in Chapters 3 through 5 are applied to the discrete shallow water equations. But first, the model error sensitivity equations from Chapter 5 are validated for the discrete shallow water model.

### 6.3.1 Validation of the Sensitivity Equations

The gradient check for the model error sensitivity equations using the finite difference approximation (5.50) is now performed using the discrete shallow water model. The same experimental setup is used in this situation as was employed for the Lorenz model, i.e. errors of the form (5.51) are calculated for  $h = 2^{-n}$  and the logarithm of the error ratios  $\log_2(E_h/E_{h/2})$  is monitored. Table 6.1 shows these error ratios using the discrete shallow water model. Like the results for the Lorenz model, the log error ratios approach 2 for each of the sensitivity equations, agreeing with the correct convergence rate of (5.50). For several values of  $h$ , no error ratios exist for the **Q**- and **C**-sensitivity equations since those choices of  $h$  produced covariance matrices that were not positive definite.

### 6.3.2 Data Assimilation Experiments

The data assimilation approaches described in Chapters 3 through 5 are applied to the discrete shallow water equations. First, the ensemble-based w4D-Var approaches from Chapter 3 are compared to a sensitivity-based DAS where the model error covariance specification is updated according to the adaptive procedure from Chapter 5. Four assimilation systems are considered where the model error covariance matrices are specified as follows.

- (Control) Mis-specified model error covariances specified as  $\mathbf{Q} = \mathbf{I}_n$  and model bias  $\mathbf{q} = \mathbf{0}$  is considered as the status-quo and serves as the basis for comparing against the other schemes.
- (Ensemble) Use equations (3.6) and (3.8) from an ensemble size of 20. Covariance localization is performed using the fifth-order rational function of Gaspari and Cohn [37] with correlation parameter 50. The model error ensemble is

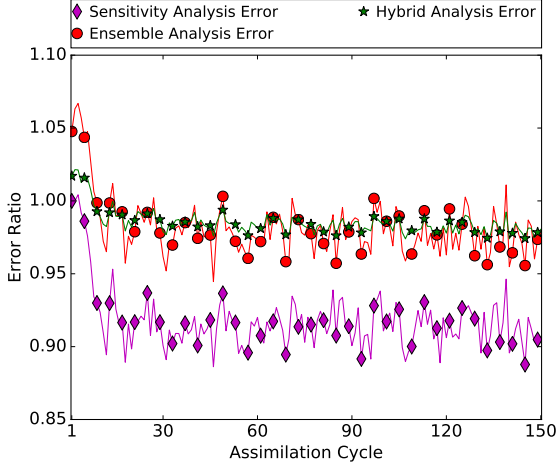


Figure 6.5: The ratio of the global analysis errors to the control w4D-Var experiment.

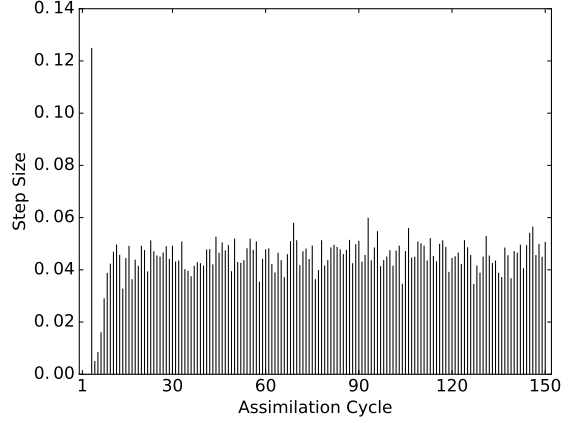


Figure 6.6: The step-sizes for the gradient descent update of the model error covariance matrices for each data assimilation cycle.

generated according to Algorithm 3.1 using the control DAS.

- (Hybrid) Use equations (3.11) and (3.12) with  $\alpha = 0.5$ . The static components are set to  $\mathbf{Q}_c = \mathbf{I}_n$  and  $\mathbf{q}_c = \mathbf{0}$  and the ensemble covariance with localization is the same one computed for the ensemble scheme.
- (Sensitivity) Initially set  $\mathbf{Q} = \mathbf{I}_n$  and the adaptive step-size to 1 for the first assimilation cycle. Then,  $\mathbf{Q}$  is updated using the gradient descent algorithm, Algorithm 5.1, via sensitivities to the square root factor  $\mathbf{Q}^{1/2}$ .

These assimilation schemes were run over a period of 150 cycles. The ratios of the average analysis errors in each cycle to the control are shown in Figure 6.5. As before, an error ratio smaller than 1 indicates better performance than the control and greater than 1 means a worse performance. For the majority of the assimilation period, the ensemble-based and sensitivity-based DASs all outperformed the control. The most significant improvement was achieved by the sensitivity-based DAS. During the assimilation period, the model error covariances from the sensitivity-based DAS were updated 147 times, for a total of 98% success rate in forecast error reduction.

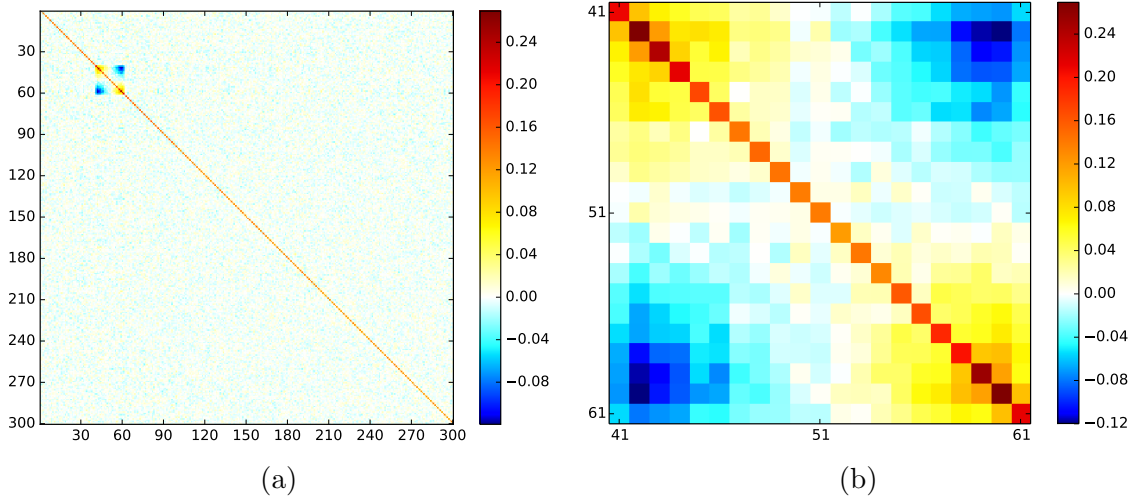


Figure 6.7: (a) Posterior estimate of  $\mathbf{Q}^t$ , (b) Zoomed-in view of the  $\mathbf{Q}$ -diagnostic matrix.

The adaptive step-sizes are shown in Figure 6.6.

Unlike the numerical results with the Lorenz model in Chapter 3, the w4D-Var ensemble DAS has outperformed the hybrid DAS. By selecting the hybrid scalar weight  $\alpha = 0.5$  in defining  $\mathbf{Q}$  by (3.11), the static part  $\mathbf{Q}_c$  plays a large role in the analysis quality. Of particular note is the fact that  $\mathbf{Q}_c = \mathbf{I}_n$  differs significantly from  $\mathbf{Q}^t$ . The entries of  $\mathbf{Q}^t$  have smaller magnitude than those of  $\mathbf{Q}_c$  and the diagonal of  $\mathbf{Q}^t$  is mostly zero. This discrepancy between  $\mathbf{Q}_c$  and  $\mathbf{Q}^t$  can explain the change in performance that was seen when compared to the Lorenz model results.

The application of the model error diagnostics of Chapter 4 to the discrete shallow water model is now considered. The specification of the control DAS provides the basis for these experiments. The  $\mathbf{Q}$ - and  $\mathbf{R}$ -diagnostic matrices are analyzed from observation residuals from 150 assimilation cycles and are calculated from (4.32) and (4.33). More specifically, the three  $\mathbf{Q}$ -diagnostic blocks of (4.32) are averaged together to estimate  $\mathbf{Q}^t$ , just as was done for the Lorenz model experiments. Similarly, the four  $\mathbf{R}$ -diagnostic blocks of (4.33) are averaged together and compared against the time-invariant  $\mathbf{R}_k^t = \mathbf{I}_n$ . The average  $\mathbf{Q}$ -diagnostic matrix for this period is shown in

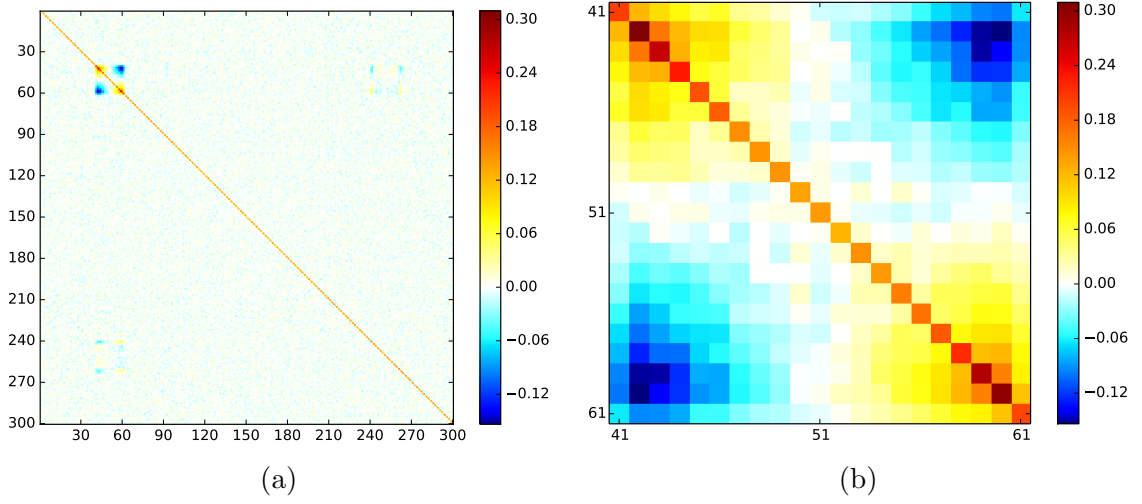


Figure 6.8: (a) Posterior estimate of  $\mathbf{R}^t$ , (b) Zoomed-in view of the  $\mathbf{R}$ -diagnostic matrix.

Figure 6.7 while the average  $\mathbf{R}$ -diagnostic matrix is shown in Figure 6.8.

Just as with the results obtained from data assimilation experiments using the Lorenz model, the  $\mathbf{Q}$ -diagnostic matrix exhibits correlations of  $\mathbf{Q}^t$  that were missing from the DAS. Unlike  $\mathbf{Q}^t$  though, which is all zero except for a  $20 \times 20$  diagonal block, the entire diagonal of the  $\mathbf{Q}$ -diagnostic matrix is nonzero. It appears that the  $\mathbf{Q}$ -diagnostic matrix has incorporated elements of both the true model error covariance matrix  $\mathbf{Q}^t$  and the DAS specification  $\mathbf{Q}$ . In any case, the  $\mathbf{Q}$ -diagnostic matrix has revealed information on the correlation structure of  $\mathbf{Q}^t$ , which was absent from the DAS, even if it is not a perfect representation of the true model error covariance matrix. Also, the  $\mathbf{R}$ -diagnostic matrix has correlations like  $\mathbf{Q}^t$ . Even though  $\mathbf{R}_k^t = \mathbf{I}_n$ , the true model error correlation structure has been incorporated into the observation covariance diagnostic matrix. This type of behavior is consistent with previous results by Todling [76] and the results from Chapter 4.



## 6.4 Conclusion

The shallow water equations provided the basis for implementing the methodologies developed in the previous chapters to establish a proof-of-concept using a model more sophisticated than the Lorenz equations. The presented results are obtained from a similar experimental design, i.e. only the model error covariance  $\mathbf{Q}$  is mis-specified in the DAS. Recovered knowledge of the model error statistics and improvements made to the analysis are the result of the successful application of the new techniques.

Through the use of model error ensembles and adaptive sensitivity-based updates, the reduction of analysis error has been achieved. The w4D-Var  $\mathbf{Q}$ -diagnostic matrix provided insight into the true model error correlation structures that were left out of the DAS, just like the results from Chapter 4 for the Lorenz model. The  $\mathbf{R}$ -diagnostic matrix displays an apparent observation correlation structure that, in truth, does not exist, replicating the behavior seen with the Lorenz model in Chapter 4 and in Todling’s results [76]. Encouraging results with this more sophisticated model have set the stage for their application in large-scale practical applications, where improved model error specification is tantamount to quality analyses.

## Chapter 7

### Conclusions and Future Directions

Accurate specification of the model error parameters in w4D-Var is linked to quality data assimilation system performance. Previously, the practical implementation of w4D-Var was dependent on assumptions about the model error in order to reduce the computational cost. The research contained in this document adds to the practical techniques and methodologies for improving the representation of the model error covariance in w4D-Var. There are still many additional avenues of research concerning model error representation in w4D-Var. In particular, further advances can be made to the ensemble- and sensitivity-based techniques presented here. Some possible research directions are outlined next.

Improvements to the ensemble-based approaches can be made by optimal selection of the ensemble size, localization operator, and hybrid covariance parameter. For a data assimilation window  $[t_0, t_N]$  of  $N + 1$  times,  $N$  model error ensembles of size  $N_e$  are formed and are used to estimate each model error covariance  $\mathbf{Q}_1, \dots, \mathbf{Q}_N$ . When the dimension of the state space is large, this can be computationally expensive, so it would be desirable to have a small ensemble size and still obtain a good estimate of the model error statistics. To do this becomes a question of how to optimally perturb the background and observations when forming the initial ensemble. Another improvement can be made by determining how to specify the hybrid scalar weights to get the best improvement in the quality of the analysis. A worthwhile research

direction is the incorporation of forecast error sensitivities into a hybrid w4D-Var DAS so that the sensitivity to the hybrid scalar weights can be developed. The sensitivity feedback mechanism of Chapter 5 should provide the foundation for tuning the hybrid scalar weights.

Further research needs to be conducted to investigate the w4D-Var diagnostics in suboptimal experiments. In particular, disentangling the model error correlations from the observation error covariance diagnostic is of interest. Since numerical experiments focused only on mis-specifying the model error covariance, the role of  $\mathbf{B}$  is not explored. However, the equations derived in Chapter 4 provide the ability to analyze  $\mathbf{B}$  as well. It remains to be seen how the diagnostics will perform when all of the error covariances are mis-specified or different observing systems are considered.

Numerical results have shown the merits of sensitivity-based tuning of the model error covariance. Future research directions include the development and refinement of algorithms for derivative-based error covariance tuning and testing with models of increased complexity.

This dissertation extends the field of data assimilation with its theoretical foundations to improve model error representation in w4D-Var. Weak-constraint 4D-Var itself is not new, but has benefited from new methodologies and algorithms this work presents. These contributions to the theory of data assimilation have led to many new problems and areas of potential research.

## References

- [1] Santha Akella and Ionel M. Navon. Different approaches to model error formulation in 4D-Var: A study with high-resolution advection schemes. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 61 A(1):112–128, 2009. 58
- [2] Brian D. O. Anderson and John B. Moore. *Optimal Filtering*. Prentice-Hall, 1979. 4
- [3] Jeffrey L. Anderson. A local least squares framework for ensemble filtering. *Monthly Weather Review*, 131(4):634–642, apr 2003. 5
- [4] Nancy L. Baker and Roger Daley. Observation and background adjoint sensitivity in the adaptive observation-targeting problem. *Quarterly Journal of the Royal Meteorological Society*, 126(565):1431–1454, jul 2007. 58
- [5] Peter J. Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36(1):199–227, feb 2008. 32
- [6] François Bouttier and Philippe Courtier. Data assimilation concepts and methods March 1999, 1999. 1
- [7] Mark Buehner. Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quarterly Journal of the Royal Meteorological Society*, 131(607):1013–1043, apr 2005. 103
- [8] Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review*, 126(6):1719–1724, jun 1998. 31
- [9] Bernard Chapnik, Gerald Desroziers, Florence Rabier, and Olivier Talagrand. Diagnosis and tuning of observational error in a quasi-operational data assimilation setting. *Quarterly Journal of the Royal Meteorological Society*, 132(615):543–565, jan 2006. 45
- [10] Alexandru Cioaca and Adrian Sandu. An optimization framework to improve 4D-Var data assimilation system performance. *Journal of Computational Physics*, 275:377–389, oct 2014. 58

- [11] Alexandru Cioaca, Adrian Sandu, and Eric de Sturler. Efficient methods for computing observation impact in 4D-Var data assimilation. *Computational Geosciences*, 17(6):975–990, dec 2013. 58
- [12] Adam M. Clayton, Andrew C. Lorenc, and Dale M. Barker. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quarterly Journal of the Royal Meteorological Society*, 139(675):1445–1461, jul 2013. 5
- [13] Philippe Courtier, Jean-Noél Thépaut, and Anthony Hollingsworth. A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519):1367–1387, jul 1994. 19
- [14] Dacian N. Daescu. On the Sensitivity Equations of Four-Dimensional Variational (4D-Var) Data Assimilation. *Monthly Weather Review*, 136(8):3050–3065, aug 2008. 58
- [15] Dacian N. Daescu. Forecast sensitivity to the observation error covariance in variational data assimilation. *Procedia Computer Science*, 1(1):1277–1285, may 2010. 58
- [16] Dacian N. Daescu and Rolf H. Langland. Error covariance sensitivity and impact estimation with adjoint 4D-Var: theoretical aspects and first applications to NAVDAS-AR. *Quarterly Journal of the Royal Meteorological Society*, 139(670):226–241, jan 2013. 58
- [17] Dacian N. Daescu and Rolf H. Langland. The Adjoint Sensitivity Guidance to Diagnosis and Tuning of Error Covariance Parameters. In Ki Seon Park and Liang Xu, editors, *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II)*, pages 205–232. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. 59
- [18] Dacian N. Daescu and Ionel M. Navon. Sensitivity Analysis in Nonlinear Variational Data Assimilation: Theoretical Aspects and Applications. In Istvan Farago and Zahari Zlatev, editors, *Advanced Numerical Methods for Complex Environmental Models: Needs and Availability*, pages 276–300. Bentham Science Publishers, 2013. 59
- [19] Dacian N. Daescu and Ricardo Todling. Adjoint Estimation of the Variation in Model Functional Output due to the Assimilation of Data. *Monthly Weather Review*, 137(5):1705–1716, may 2009. 58
- [20] Dacian N. Daescu and Ricardo Todling. Adjoint sensitivity of the model forecast to data assimilation system error covariance parameters. *Quarterly Journal of the Royal Meteorological Society*, 136(653):2000–2012, oct 2010. 58

- [21] Dick P. Dee. Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3323–3343, oct 2005. 29, 45
- [22] Dick P. Dee and Arlindo M. Da Silva. Data assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society*, 124(545):269–295, jan 1998. 29, 45
- [23] Dick P. Dee and Ricardo Todling. Data Assimilation in the Presence of Forecast Bias: The GEOS Moisture Analysis. *Monthly Weather Review*, 128(9):3268–3282, sep 2000. 29
- [24] Dick P. Dee and Sakari Uppala. Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 135(644):1830–1841, 2009. 29
- [25] John C. Derber. A variational continuous assimilation technique. *Monthly Weather Review*, 117(11):2437–2446, 1989. 7
- [26] Gérald Desroziers, Loïk Berre, Vincent Chabot, and Bernard Chapnik. A Posteriori Diagnostics in an Ensemble of Perturbed Analyses. *Monthly Weather Review*, 137(10):3420–3436, oct 2009. 58
- [27] Gérald Desroziers, Loïk Berre, Bernard Chapnik, and Paul Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3385–3396, oct 2005. 45, 48, 49, 50, 51, 56
- [28] Gérald Desroziers and Serguei Ivanov. Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, 127(574):1433–1452, apr 2001. 45, 58
- [29] Jean Philippe Drécourt, Henrik Madsen, and Dan Rosbjerg. Bias aware Kalman filters: Comparison and improvements. *Advances in Water Resources*, 29:707–718, 2006. 29
- [30] Bradley Efron and Carl Morris. Multivariate Empirical Bayes and Estimation of Covariance Matrices. *The Annals of Statistics*, 4(1):22–32, 1976. 34
- [31] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5):10143, 1994. 4, 29
- [32] Geir Evensen. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, 54(6):539–560, dec 2004. 5
- [33] Geir Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer-Verlag Berlin Heidelberg, 2nd edition, 2009. 4

- [34] Geir Evensen and Peter Jan van Leeuwen. An Ensemble Kalman Smoother for Nonlinear Dynamics. *Monthly Weather Review*, 128(6):1852–1867, jun 2000. 5
- [35] David Fairbairn, Stephen R. Pring, Andrew C. Lorenc, and Ian Roulstone. A comparison of 4DVar with ensemble data assimilation methods. *Quarterly Journal of the Royal Meteorological Society*, 140(678):281–294, jan 2014. 29
- [36] Mike Fisher, Yannick Trémolet, Harri Auvinen, David Tan, and Paul Poli. Weak-constraint and long window 4DVAR, nov 2011. 16
- [37] Gregory Gaspari and Stephen E. Cohn. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757, jan 1999. 34, 88
- [38] Ralf Giering and Thomas Kaminski. Recipes for adjoint code construction. *ACM Transactions on Mathematical Software*, 24(4):437–474, dec 1998. 8
- [39] Mohinder Grewal and Angus Andrews. Applications of Kalman Filtering in Aerospace 1960 to the Present [Historical Perspectives. *IEEE Control Systems Magazine*, 30(3):69–78, jun 2010. 3
- [40] Andreas Griewank and Andrea Walther. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, second edition, 2008. 8
- [41] Anne K. Griffith and Nancy K. Nichols. Adjoint Methods in Data Assimilation for Estimating Model Error. *Flow, Turbulence and Combustion*, 65(3/4):469–488, 2000. 7, 82
- [42] Thomas M. Hamill and Chris Snyder. A Hybrid Ensemble Kalman Filter3D Variational Analysis Scheme. *Monthly Weather Review*, 128(8):2905–2919, aug 2000. 5
- [43] Thomas M. Hamill, Jeffrey S. Whitaker, and Chris Snyder. Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. *Monthly Weather Review*, 129(11):2776–2790, nov 2001. 5, 29
- [44] Laurent Hascoet and Valérie Pascual. The Tapenade Automatic Differentiation Tool: Principles, Model, and Specification. *ACM Trans. Math. Softw.*, 39(3):20:1–20:43, 2013. 9, 108
- [45] Anthony Hollingsworth and Peter Lönnberg. The verification of objective analyses: Diagnostics of analysis system performance. *Meteorology and Atmospheric Physics*, 40(1-3):3–27, 1989. 45
- [46] Daisuke Hotta. *Proactive Quality Control Based on Ensemble Forecast*. PhD thesis, University of Maryland, College Park, 2014. 77

- [47] Peter L. Houtekamer and Herschel L. Mitchell. A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, 129(1):123–137, jan 2001. 5
- [48] Peter L. Houtekamer and Herschel L. Mitchell. Ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 131:3269–3289, 2005. 29
- [49] Peter L. Houtekamer, Herschel L. Mitchell, Gérard Pellerin, Mark Buehner, Martin Charron, Lubos Spacek, and Bjarne Hansen. Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations. *Monthly Weather Review*, 133(3):604–620, 2005. 29
- [50] Andrew H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970. 4, 18
- [51] Rudolf E. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(Series D):35–45, 1960. 2
- [52] Rudolf E. Kalman and Richard S. Bucy. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83(1):95–108, 1961. 2
- [53] William Lahoz, Boris Khattatov, and Richard Ménard, editors. *Data Assimilation: Making Sense of Observations*. Springer-Verlag Berlin Heidelberg, 2010. 1
- [54] Francois-Xavier Le Dimet, Ionel M. Navon, and Dacian N. Daescu. Second-Order Information in Data Assimilation. *Monthly Weather Review*, 130(3):629–648, mar 2002. 62
- [55] Francois-Xavier Le Dimet, Hans E. Ngodock, B. Luong, and J. Verron. Sensitivity analysis in variational data assimilation. *Journal of the Meteorological Society of Japan*, 75(1B):245–255, 1997. 58
- [56] Francois-Xavier Le Dimet and Olivier Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A*, 38A(2):97–110, mar 1986. 5
- [57] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411, feb 2004. 35
- [58] Hongze Leng, Junqiang Song, Fengshun Lu, and Xiaoqun Cao. A New Data Assimilation Scheme: The Space-Expanded Ensemble Localization Kalman Filter. *Advances in Meteorology*, 2013:1–6, 2013. 103



- [59] John M. Lewis, S. Lakshmivarahan, and Sudarshan Dhall. *Dynamic Data Assimilation: A Least Squares Approach (Encyclopedia of Mathematics and its Applications)*. Cambridge University Press, 2006. 16
- [60] Andrew C. Lorenc. Optimal nonlinear objective analysis. *Quarterly Journal of the Royal Meteorological Society*, 114(479):205–240, jan 1988. 82
- [61] Andrew C. Lorenc. The potential of the ensemble Kalman filter for NWP: a comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 129(595):3183–3203, oct 2003. 29
- [62] Andrew C. Lorenc and Richard T. Marriott. Forecast sensitivity to observations in the Met Office Global numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, 140(678):209–224, jan 2014. 59
- [63] Edward N. Lorenz. Predictability: A Problem Partly Solved. In *Seminar on Predictability, 4-8 September 1995*, volume 1, pages 1–18, Shinfield Park, Reading, 1995. ECMWF, ECMWF. 10
- [64] Edward N. Lorenz and Kerry A. Emanuel. Optimal Sites for Supplementary Weather Observations: Simulation with a Small Model. *Journal of the Atmospheric Sciences*, 55(3):399–414, feb 1998. 11
- [65] Cristina Lupu, Carla Cardinali, and Anthony P. McNally. Adjoint-based forecast sensitivity applied to observation-error variance tuning. *Quarterly Journal of the Royal Meteorological Society*, 141(693):3157–3165, oct 2015. 58
- [66] Herschel L. Mitchell and Peter L. Houtekamer. An Adaptive Ensemble Kalman Filter. *Monthly Weather Review*, 128(2):416, 2000. 5
- [67] Nancy K. Nichols. Mathematical Concepts of Data Assimilation. In *Data Assimilation*, pages 13–39. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. 1
- [68] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2006. 71
- [69] Margarida Belo Pereira and Loïk Berre. The Use of an Ensemble Approach to Study the Background Error Covariances in a Global NWP Model. *Monthly Weather Review*, 134(9):2466–2489, sep 2006. 31
- [70] Thomas E. Rosmond and Liang Xu. Development of NAVDAS-AR: non-linear formulation and outer loop tests. *Tellus A*, 58(1):45–58, jan 2006. 28
- [71] Yoshikazu Sasaki. Some Basic Formalisms in Numerical Variational Analysis. *Monthly Weather Review*, 98(12):875–883, dec 1970. 6

- [72] Jeremy A. Shaw and Dacian N. Daescu. An Ensemble Approach to Weak-constraint Four-dimensional Variational Data Assimilation. *Procedia Computer Science*, 80:496–506, 2016. 15, 30
- [73] Hajoon Song, Ibrahim Hoteit, Bruce D. Cornuelle, Xiaodong Luo, and Aneesh C. Subramanian. An Adjoint-Based Adaptive Ensemble Kalman Filter. *Monthly Weather Review*, 141(10):3343–3359, oct 2013. 69
- [74] Olivier Talagrand. Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan*, 75(1B):191–209, oct 1997. 1
- [75] Olivier Talagrand and Philippe Courtier. Variational Assimilation of Meteorological Observations With the Adjoint Vorticity Equation. I: Theory. *Quarterly Journal of the Royal Meteorological Society*, 113(478):1311–1328, jul 2007. 6
- [76] Ricardo Todling. A complementary note to A lag-1 smoother approach to system-error estimation’: the intrinsic limitations of residual diagnostics. *Quarterly Journal of the Royal Meteorological Society*, 141(692):2917–2922, oct 2015. 45, 56, 91, 92
- [77] Ricardo Todling. A lag-1 smoother approach to system-error estimation: sequential method. *Quarterly Journal of the Royal Meteorological Society*, 141(690):1502–1513, jul 2015. 45
- [78] Yannick Trémolet. Accounting for an imperfect model in 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 132(621):2483–2504, oct 2006. 19
- [79] Jeffrey S. Whitaker and Thomas M. Hamill. Ensemble Data Assimilation without Perturbed Observations. *Monthly Weather Review*, 130(7):1913–1924, jul 2002. 31
- [80] Christopher K. Wikle and L. Mark Berliner. A Bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena*, 230(1-2):1–16, jun 2007. 18
- [81] Max Yaremchuk and Paul Martin. On Sensitivity Analysis within the 4DVAR Framework. *Monthly Weather Review*, 142(2):774–787, feb 2014. 58
- [82] Dusanka Zupanski. A General Weak Constraint Applicable to Operational 4DVAR Data Assimilation Systems. *Monthly Weather Review*, 125(9):2274–2292, sep 1997. 7
- [83] Dusanka Zupanski and Milija Zupanski. Model Error Estimation Employing an Ensemble Data Assimilation Approach. *Monthly Weather Review*, 134(5):1337–1354, 2006. 7

## Appendix A

### Ensemble Covariance Implementation

Since it is necessary to compute the product of the model error covariance with a vector to obtain the analysis, this appendix provides notes on the implementation of  $\mathbf{Q}_i^{1/2}$  when  $\mathbf{Q}_i$  is the ensemble error covariance with localization or the hybrid error covariance. The ensemble error covariance (3.7) can be expressed by

$$\mathbf{Q}_{i,e} = \sum_{j=1}^{N_e} \mathbf{q}_{i,j} \mathbf{q}_{i,j}^T, \quad (\text{A.1})$$

by defining

$$\mathbf{q}_{i,j} = \frac{1}{\sqrt{N_e - 1}} (\boldsymbol{\eta}_{i,j} - \mathbf{q}_{i,e}). \quad (\text{A.2})$$

From (A.1), the product of a vector  $\mathbf{x} \in \mathbb{R}^n$  with  $\mathbf{Q}_{i,e}$  can be calculated as follows.

$$\mathbf{Q}_{i,e} \mathbf{x} = \sum_{j=1}^{N_e} (\mathbf{q}_{i,j}^T \mathbf{x}) \mathbf{q}_{i,j} \quad (\text{A.3})$$

When a localization matrix  $\mathbf{C}_i$  is applied to the ensemble error covariance matrix, the product with  $\mathbf{x} \in \mathbb{R}^n$  is

$$(\mathbf{Q}_{i,e} \circ \mathbf{C}_i) \mathbf{x} = \sum_{j=1}^{N_e} (\mathbf{C}_i (\mathbf{x} \circ \mathbf{q}_{i,j})) \circ \mathbf{q}_{i,j}. \quad (\text{A.4})$$

Since it can be shown that

$$(\mathbf{q}_{i,j} \mathbf{q}_{i,j}^T) \circ \mathbf{C}_i = \text{diag}(\mathbf{q}_{i,j}) \mathbf{C}_i \text{diag}(\mathbf{q}_{i,j}), \quad (\text{A.5})$$

it follows that  $\mathbf{Q}_i = \mathbf{Q}_{i,e} \circ \mathbf{C}_i$  can be factored in the form  $\mathbf{Q}_i = \mathbf{Q}_i^{1/2} \mathbf{Q}_i^{T/2}$  by letting

$$\mathbf{Q}_i^{1/2} = \begin{bmatrix} \text{diag}(\mathbf{q}_{i,1}) \mathbf{C}_i^{1/2} & \cdots & \text{diag}(\mathbf{q}_{i,N_e}) \mathbf{C}_i^{1/2} \end{bmatrix} \in \mathbb{R}^{n \times (nN_e)}. \quad (\text{A.6})$$

Other studies have implemented this decomposition for ensemble-based data assimilation, such as [7] and [58]. Multiplication of  $\mathbf{Q}_i^{1/2}$  with a vector  $\mathbf{x} \in \mathbb{R}^{nN_e}$ , which can be composed into  $N_e$  blocks of size  $n$  according to

$$\mathbf{x} = [\mathbf{x}_1^T \quad \cdots \quad \mathbf{x}_{N_e}^T]^T, \quad (\text{A.7})$$

can be evaluated by

$$\mathbf{Q}_i^{1/2} \mathbf{x} = \sum_{j=1}^{N_e} (\mathbf{C}_i^{1/2} \mathbf{x}_j) \circ \mathbf{q}_{i,j} \in \mathbb{R}^n. \quad (\text{A.8})$$

The product of  $\mathbf{Q}_i^{T/2}$  with a vector  $\mathbf{x} \in \mathbb{R}^n$  is

$$\mathbf{Q}_i^{T/2} \mathbf{x} = \left[ \mathbf{C}_i^{T/2} (\mathbf{q}_{i,j} \circ \mathbf{x}) \right]_{j=1}^{N_e} \in \mathbb{R}^{nN_e}. \quad (\text{A.9})$$

The hybrid covariance (3.10) can be implemented in a similar way. The square root of the hybrid error covariance

$$\mathbf{Q}_i^{1/2} = \left[ \sqrt{1 - \alpha_i} \text{diag}(\mathbf{q}_{i,1}) \mathbf{C}_i^{1/2} \quad \cdots \quad \sqrt{1 - \alpha_i} \text{diag}(\mathbf{q}_{i,N_e}) \mathbf{C}_i^{1/2} \quad \sqrt{\alpha_i} \mathbf{Q}_{i,c}^{1/2} \right] \quad (\text{A.10})$$

has dimension  $n \times n(N_e + 1)$ . For  $\mathbf{x} \in \mathbb{R}^{n(N_e+1)}$ , which can be composed into  $N_e + 1$  blocks of size  $n$  according to

$$\mathbf{x} = [\mathbf{x}_1^T \quad \cdots \quad \mathbf{x}_{N_e}^T \quad \mathbf{x}_{N_e+1}^T]^T, \quad (\text{A.11})$$

the product of  $\mathbf{Q}_i^{1/2}$  with  $\mathbf{x}$  is

$$\mathbf{Q}_i^{1/2} \mathbf{x} = \sqrt{1 - \alpha_i} \sum_{j=1}^{N_e} (\mathbf{C}_i^{1/2} \mathbf{x}_j) \circ \mathbf{q}_{i,j} + \sqrt{\alpha_i} \mathbf{Q}_{i,c}^{1/2} \mathbf{x}_{N_e+1} \in \mathbb{R}^n. \quad (\text{A.12})$$

If  $\mathbf{y} = \mathbf{Q}_i^{T/2} \mathbf{x}$  where  $\mathbf{x} \in \mathbb{R}^n$ , then  $\mathbf{y}$  can be decomposed into  $N_e + 1$  blocks of size  $n$

$$\mathbf{y} = [\mathbf{y}_1^T \quad \cdots \quad \mathbf{y}_{N_e}^T \quad \mathbf{y}_{N_e+1}^T]^T, \quad (\text{A.13})$$

where

$$\mathbf{y}_j = \sqrt{1 - \alpha_i} \mathbf{C}_i^{T/2} (\mathbf{q}_{i,j} \circ \mathbf{x}), \quad j = 1, \dots, N_e, \quad \mathbf{y}_{N_e+1} = \sqrt{\alpha_i} \mathbf{Q}_{i,c}^{T/2} \mathbf{x}. \quad (\text{A.14})$$

## Appendix B

### The Discrete Shallow Water Equations

The shallow water equations (6.1) can be rewritten in flux form by replacing (6.1a) with  $\phi \times (6.1a) + u \times (6.1c)$  and by replacing (6.1b) with  $\phi \times (6.1b) + v \times (6.1c)$ .

$$\frac{\partial}{\partial t} [\phi u] + \frac{\partial}{\partial x} \left[ \phi u^2 + \frac{1}{2} \phi^2 \right] = f \phi v - g \phi \frac{\partial H}{\partial x} \quad (\text{B.1a})$$

$$\frac{\partial}{\partial t} [\phi v] + \frac{\partial}{\partial x} [\phi uv] = -f \phi u \quad (\text{B.1b})$$

$$\frac{\partial \phi}{\partial t} + \frac{\partial}{\partial x} [\phi u] = 0 \quad (\text{B.1c})$$

If  $m = \phi u$  and  $n = \phi v$ , the shallow water equations can be rewritten as follows.

$$\frac{\partial m}{\partial t} + \frac{\partial}{\partial x} \left[ mu + \frac{1}{2} \phi^2 \right] = f n - g \phi \frac{\partial H}{\partial x} \quad (\text{B.2a})$$

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial x} [mv] = -f m \quad (\text{B.2b})$$

$$\frac{\partial \phi}{\partial t} + \frac{\partial m}{\partial x} = 0 \quad (\text{B.2c})$$

Diffusion terms are added to each subequation of (B.2) to help eliminate spurious oscillations in the discrete shallow water equations.

$$\frac{\partial m}{\partial t} + \frac{\partial}{\partial x} \left[ mu + \frac{1}{2} \phi^2 \right] = f n - g \phi \frac{\partial H}{\partial x} + K \frac{\partial^2 m}{\partial x^2} \quad (\text{B.3a})$$

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial x} [mv] = -f m + K \frac{\partial^2 n}{\partial x^2} \quad (\text{B.3b})$$

$$\frac{\partial \phi}{\partial t} + \frac{\partial m}{\partial x} = K \frac{\partial^2 \phi}{\partial x^2} \quad (\text{B.3c})$$

The shallow water equations (B.3) are discretized using forward time and centered space finite differences. The time derivatives are estimated using the first-order

accurate forward difference. Thus

$$\frac{\partial m}{\partial t} = \frac{m_j^{k+1} - m_j^k}{\Delta t} + \mathcal{O}(\Delta t) \quad (\text{B.4})$$

and similar expressions are used for the time derivatives of  $n(x, t)$  and  $\phi(x, t)$ .

The product rule of differentiation implies that

$$\frac{\partial}{\partial x} [mu] = m \frac{\partial u}{\partial x} + u \frac{\partial m}{\partial x}, \quad (\text{B.5})$$

and each side of this equation can be estimated using centered differencing

$$\frac{\partial}{\partial x} [mu] = \frac{m_{j+1}^k u_{j+1}^k - m_{j-1}^k u_{j-1}^k}{2\Delta x} + \mathcal{O}(\Delta x^2) \quad (\text{B.6})$$

$$m \frac{\partial u}{\partial x} + u \frac{\partial m}{\partial x} = m_j^k \frac{u_{j+1}^k - u_{j-1}^k}{2\Delta x} + u_j^k \frac{m_{j+1}^k - m_{j-1}^k}{2\Delta x} + \mathcal{O}(\Delta x^2). \quad (\text{B.7})$$

The average of these two schemes is also a second-order accurate approximation of  $\partial(mu)/\partial x$ . After adding and subtracting the term  $m_j^k u_j^k$  to  $\frac{1}{2}[(\text{B.6}) + (\text{B.7})]$  and some factoring, we get the finite difference approximation for  $\partial(mu)/\partial x$ .

$$\begin{aligned} \frac{\partial}{\partial x} [mu] &= \frac{1}{4\Delta x} [(u_{j+1}^k + u_j^k)(m_{j+1}^k + m_j^k) \\ &\quad - (u_j^k + u_{j-1}^k)(m_j^k + m_{j-1}^k)] + \mathcal{O}(\Delta x^2) \end{aligned} \quad (\text{B.8})$$

A similar process yields the finite difference approximation for  $\partial(mv)/\partial x$ .

$$\begin{aligned} \frac{\partial}{\partial x} [mv] &= \frac{1}{4\Delta x} [(v_{j+1}^k + v_j^k)(m_{j+1}^k + m_j^k) \\ &\quad - (v_j^k + v_{j-1}^k)(m_j^k + m_{j-1}^k)] + \mathcal{O}(\Delta x^2) \end{aligned} \quad (\text{B.9})$$

Next, the spatial derivative of  $\phi^2$  is simply the standard centered difference formula.

$$\frac{\partial}{\partial x} [\phi^2] = \frac{(\phi_{j+1}^k)^2 - (\phi_{j-1}^k)^2}{2\Delta x} + \mathcal{O}(\Delta x^2) \quad (\text{B.10})$$

On the right side of equation (B.3a), there is  $\phi \partial H / \partial x$ , which can be written as

$$\phi \frac{\partial H}{\partial x} = \frac{\partial}{\partial x} [\phi H] - H \frac{\partial \phi}{\partial x} \quad (\text{B.11})$$

from the product rule of differentiation. In a manner similar to the finite difference

scheme from (B.8), we get the following.

$$\frac{\partial}{\partial x} \left[ \phi H \right] - H \frac{\partial \phi}{\partial x} = \frac{\phi_{j+1}^k H_{j+1} - \phi_{j-1}^k H_{j-1}}{2\Delta x} - H_j \frac{\phi_{j+1}^k - \phi_{j-1}^k}{2\Delta x} + \mathcal{O}(\Delta x^2) \quad (\text{B.12})$$

$$\phi \frac{\partial H}{\partial x} = \phi_j^k \frac{H_{j+1} - H_{j-1}}{2\Delta x} + \mathcal{O}(\Delta x^2) \quad (\text{B.13})$$

The average of equations (B.12) and (B.13) gives another second-order approximation for  $\phi \partial H / \partial x$ .

$$\begin{aligned} \phi \frac{\partial H}{\partial x} = \frac{1}{4\Delta x} & \left[ (\phi_{j+1}^k + \phi_j^k)(H_{j+1} - H_j) \right. \\ & \left. + (\phi_j^k + \phi_{j-1}^k)(H_j - H_{j-1}) \right] + \mathcal{O}(\Delta x^2) \end{aligned} \quad (\text{B.14})$$

Finally, the diffusion terms are expressed using the standard second-order approximation for the second derivative. Using this and the finite differences derived above, the discretization of the shallow water equations (B.3) gives (6.2).

## Appendix C

### Tangent Linear and Adjoint of the Discrete Shallow Water Model

The derivation of the tangent linear and adjoint models of the discrete shallow water equations (6.4) can be obtained by using the chain rule of differentiation. Guided by (1.23), it can be verified through a pen and paper calculation, or by using an automatic differentiation tool, that the tangent linear model is given by

$$\begin{aligned} \dot{m}_j^{k+1} = & \dot{m}_j^k - \frac{\Delta t}{4\Delta x} \left[ \left( \frac{\dot{m}_{j+1}^k - \dot{\phi}_{j+1}^k u_{j+1}^k}{\phi_{j+1}^k} + \frac{\dot{m}_j^k - \dot{\phi}_j^k u_j^k}{\phi_j^k} \right) (m_{j+1}^k + m_j^k) + \right. \\ & (u_{j+1}^k + u_j^k) (\dot{m}_{j+1}^k + \dot{m}_j^k) - \left( \frac{\dot{m}_j^k - \dot{\phi}_j^k u_j^k}{\phi_j^k} + \frac{\dot{m}_{j-1}^k - \dot{\phi}_{j-1}^k u_{j-1}^k}{\phi_{j-1}^k} \right) (m_j^k + m_{j-1}^k) - \\ & \left. (u_j^k + u_{j-1}^k) (\dot{m}_j^k + \dot{m}_{j-1}^k) + 2\phi_{j+1}^k \dot{\phi}_{j+1}^k - 2\phi_{j-1}^k \dot{\phi}_{j-1}^k \right] + \Delta t f \dot{n}_j^k + \\ & K \frac{\Delta t}{\Delta x^2} [\dot{m}_{j-1}^k - 2\dot{m}_j^k + \dot{m}_{j+1}^k] \end{aligned} \quad (\text{C.1a})$$

$$\begin{aligned} \dot{n}_j^{k+1} = & \dot{n}_j^k - \frac{\Delta t}{4\Delta x} \left[ \left( \frac{\dot{n}_{j+1}^k - \dot{\phi}_{j+1}^k v_{j+1}^k}{\phi_{j+1}^k} + \frac{\dot{n}_j^k - \dot{\phi}_j^k v_j^k}{\phi_j^k} \right) (m_{j+1}^k + m_j^k) + \right. \\ & (v_{j+1}^k + v_j^k) (\dot{m}_{j+1}^k + \dot{m}_j^k) - \left( \frac{\dot{n}_j^k - \dot{\phi}_j^k v_j^k}{\phi_j^k} + \frac{\dot{n}_{j-1}^k - \dot{\phi}_{j-1}^k v_{j-1}^k}{\phi_{j-1}^k} \right) (m_j^k + m_{j-1}^k) - \\ & \left. (v_j^k + v_{j-1}^k) (\dot{m}_j^k + \dot{m}_{j-1}^k) \right] - \Delta t f \dot{n}_j^k + K \frac{\Delta t}{\Delta x^2} [\dot{n}_{j-1}^k - 2\dot{n}_j^k + \dot{n}_{j+1}^k] \end{aligned} \quad (\text{C.1b})$$

$$\dot{\phi}_j^{k+1} = \dot{\phi}_j^k - \frac{\Delta t}{2\Delta x} [\dot{m}_{j+1}^k - \dot{m}_{j-1}^k] + K \frac{\Delta t}{\Delta x^2} [\dot{\phi}_{j-1}^k - 2\dot{\phi}_j^k + \dot{\phi}_{j+1}^k], \quad (\text{C.1c})$$

where the expressions for

$$\dot{u}_j^k = \frac{\dot{m}_j^k - \dot{\phi}_j^k u_j^k}{\phi_j^k} \quad (\text{C.2})$$



and

$$\dot{v}_j^k = \frac{\dot{n}_j^k - \phi_j^k v_j^k}{\phi_j^k} \quad (\text{C.3})$$

are used, coming from the definitions of  $m_j^k = \phi_j^k u_j^k$  and  $n_j^k = \phi_j^k v_j^k$ .

By following (1.24) and using the derivative formulae

$$\frac{\partial u_j^k}{\partial m_j^k} = \frac{\partial v_j^k}{\partial n_j^k} = \frac{1}{\phi_j^k}, \quad \frac{\partial u_j^k}{\partial \phi_j^k} = -\frac{u_j^k}{\phi_j^k}, \quad \frac{\partial v_j^k}{\partial \phi_j^k} = -\frac{v_j^k}{\phi_j^k} \quad (\text{C.4})$$

as needed, the derivation of the adjoint of (6.4) gives the following.

$$\begin{aligned} \bar{m}_j^k &= \bar{m}_j^{k+1} - \frac{\Delta t}{4\Delta x} \left[ \left( \frac{m_{j+1}^k + m_j^k}{\phi_j^k} + u_{j+1}^k + u_j^k \right) (\bar{m}_j^{k+1} - \bar{m}_{j+1}^{k+1}) + \right. \\ &\quad \left. \left( \frac{m_j^k + m_{j-1}^k}{\phi_j^k} + u_j^k + u_{j-1}^k \right) (\bar{m}_{j-1}^{k+1} - \bar{m}_j^{k+1}) \right] + \\ &\quad \frac{\Delta t}{4\Delta x} \left[ (v_{j+1}^k + v_j^k) (\bar{n}_{j+1}^{k+1} - \bar{n}_j^{k+1}) + (v_j^k + v_{j-1}^k) (\bar{n}_j^{k+1} - \bar{n}_{j-1}^{k+1}) \right] + \\ &\quad \frac{\Delta t}{2\Delta x} [\bar{\phi}_{j+1}^{k+1} - \bar{\phi}_{j-1}^{k+1}] - \Delta t f \bar{n}_j^{k+1} + K \frac{\Delta t}{\Delta x^2} [\bar{m}_{j-1}^{k+1} - 2\bar{m}_j^{k+1} + \bar{m}_{j+1}^{k+1}] \end{aligned} \quad (\text{C.5a})$$

$$\begin{aligned} \bar{n}_j^k &= \bar{n}_j^{k+1} - \frac{\Delta t}{4\Delta x \phi_j^k} \left[ (m_{j+1}^k + m_j^k) (\bar{n}_j^{k+1} - \bar{n}_{j+1}^{k+1}) + (m_j^k + m_{j-1}^k) (\bar{n}_{j-1}^{k+1} - \bar{n}_j^{k+1}) \right] + \\ &\quad \Delta t f \bar{m}_j^{k+1} + K \frac{\Delta t}{\Delta x^2} [\bar{n}_{j-1}^{k+1} - 2\bar{n}_j^{k+1} + \bar{n}_{j+1}^{k+1}] \end{aligned} \quad (\text{C.5b})$$

$$\begin{aligned} \bar{\phi}_j^k &= \bar{\phi}_j^{k+1} - \frac{\Delta t v_j^k}{4\Delta x \phi_j^k} \left[ (m_{j+1}^k + m_j^k) (\bar{n}_{j+1}^{k+1} - \bar{n}_j^{k+1}) + (m_j^k + m_{j-1}^k) (\bar{n}_j^{k+1} - \bar{n}_{j-1}^{k+1}) \right] - \\ &\quad \frac{\Delta t u_j^k}{4\Delta x \phi_j^k} \left[ (m_{j+1}^k + m_j^k) (\bar{m}_{j+1}^{k+1} - \bar{m}_j^{k+1}) + (m_j^k + m_{j-1}^k) (\bar{m}_j^{k+1} - \bar{m}_{j-1}^{k+1}) \right] + \\ &\quad \frac{\Delta t}{2\Delta x} \phi_j^k [\bar{m}_{j+1}^{k+1} - \bar{m}_{j-1}^{k+1}] + K \frac{\Delta t}{\Delta x^2} [\bar{\phi}_{j-1}^{k+1} - 2\bar{\phi}_j^{k+1} + \bar{\phi}_{j+1}^{k+1}] \end{aligned} \quad (\text{C.5c})$$

Automatic differentiation tools, like Tapenade [44], may not present the tangent linear and adjoint codes in the same manner as presented here since some optimizations are performed that yield a better code. In some cases, these optimization may make the code somewhat harder to understand.

To ensure the accuracy of the tangent linear (C.1) and adjoint (C.5) models, it is necessary to perform some numerical tests. The tangent linear model can be validated

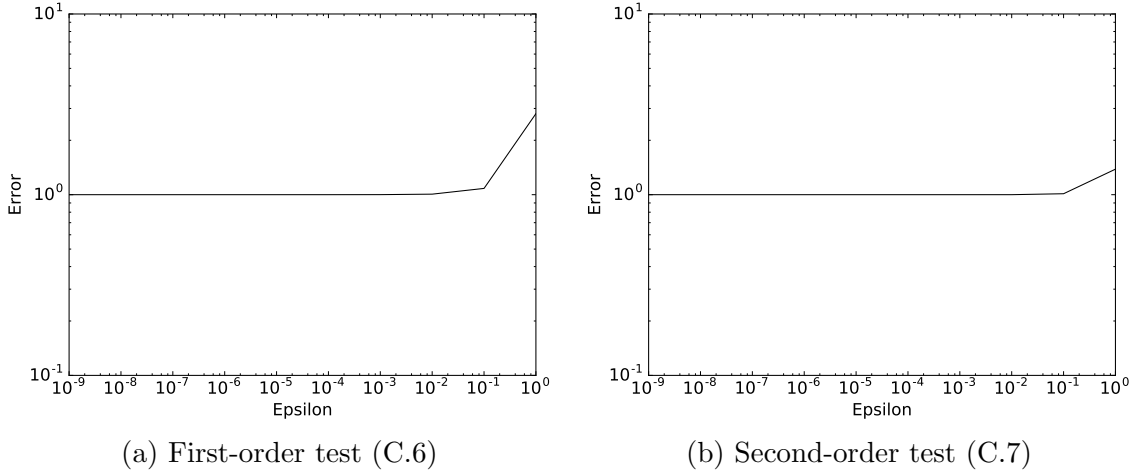


Figure C.1: Numerical validation plots for the tangent linear model.

by using the first- and second-order accurate identities

$$\frac{\|\mathcal{M}(\mathbf{x} + \varepsilon\Delta\mathbf{x}) - \mathcal{M}(\mathbf{x})\|}{\|\varepsilon\mathbf{M}\Delta\mathbf{x}\|} \approx 1 \quad (\text{C.6})$$

$$\frac{\|\mathcal{M}(\mathbf{x} + \varepsilon\Delta\mathbf{x}) - \mathcal{M}(\mathbf{x} - \varepsilon\Delta\mathbf{x})\|}{\|2\varepsilon\mathbf{M}\Delta\mathbf{x}\|} \approx 1 \quad (\text{C.7})$$

as  $\varepsilon$  approaches 0, where  $\mathbf{M}$  represents the tangent linear forecast model evaluated at a randomly generated state  $\mathbf{x}$ . For these tests, the step-size is  $\varepsilon = 10^{-i}$  for  $i = 0, 1, \dots, 9$  and the direction vector  $\Delta\mathbf{x}$  is randomly generated and normalized to unit length. The plots shown in Figure C.1 both validate the tangent linear model (C.1) as the left-hand-sides of (C.6) and (C.7) converge to 1.

Next, the adjoint model can be verified from the definition of the adjoint of a linear operator

$$\langle \mathbf{M}\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{M}^T\mathbf{v} \rangle \quad (\text{C.8})$$

by checking that the left and right sides give the same value, or equivalently,

$$\langle \mathbf{M}\mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{u}, \mathbf{M}^T\mathbf{v} \rangle = 0. \quad (\text{C.9})$$

Using randomly generated vectors  $\mathbf{u}$  and  $\mathbf{v}$ , Table C.1 shows the results of (C.9) ap-

Tangent Linear Product	Adjoint Product	Difference
-345.82026685450165360	-345.82026685450176728	$1.13686837722 \times 10^{-13}$

Table C.1: The results of the adjoint test (C.9) for the discrete shallow water model.

plied to the tangent linear and adjoint models of the discrete shallow water equations. The “tangent linear product” refers to the left side of (C.8) and “adjoint product” corresponds to the right side. The difference between the inner products is essentially zero, establishing the validity of the adjoint model (C.5). Subsequent executions of this test produce different numbers, due to randomness, but the inner product difference is always around the order of  $10^{-13}$  or smaller.